

Japan's beef scandal

The detection of BSE in Japan last week raises the prospect of vCJD to follow. The history of government handling of human health crises gives little basis for confidence that appropriate preventative measures will be taken.

Beef is a favourite food in Japan, especially among the young, with a steady yearly demand of about 1.5 million tonnes. The recent discovery of a cow suffering from bovine spongiform encephalopathy (BSE) has therefore caused considerable concern (see page 337). People are starting to dread the appearance of new variant Creutzfeldt–Jakob disease (vCJD), the human nerve-degenerating disease believed to be caused by eating certain parts of BSE-afflicted cows.

The fears are all the more palpable because the Japanese government has shown that it cannot be trusted to do the right thing in such public-health cases.

The government's complicity in covering up, and thereby spreading, mercury poisoning in Minamata during the 1950s and 1960s is the most famous example. Although researchers identified fish contaminated with industrial effluvia from a local factory as the culprit, the government took no action for years. Victims died or suffered lifelong disabilities in the following decades, and suits against the government and the company continue to this day.

The government has been slow to act in other cases as well. HIV-tainted blood was given to haemophilia patients in 1985 and 1986, long after they should have switched to heat-treated blood that was known to be safe. In 1997, the health ministry belatedly banned the import and use of contaminated German dura mater grafts, used in brain surgery, almost 10 years after the US Food and Drug Administration had taken action against them. The grafts inflicted another deadly CJD-variant on a reported 76 (and rising) brain-surgery patients, many of whom could have been saved by prompt action.

The Japanese government seems to be up to its old tricks again with BSE. A European Commission report of a survey on the risk of BSE in various countries was effectively blocked when the Japanese government insisted that it did not deserve the high-risk rating that

the commission was planning, according to commission officials. And when the existence of the mad cow was finally admitted by the government, agriculture ministry officials reportedly claimed to have burned the cow to prevent transmission — only to admit later that it had been passed on for feed production. Now they are scrambling to identify the cause, and the amount of spread in food chains.

During the early 1990s, Britain scandalously dumped much of its unwanted meat and bone meal on Japan and other Asian countries. Now there is much concern that BSE cases will spread throughout Asia, especially Indonesia and Thailand, which imported large amounts of meat and bone meal. Japan should have the economic might and regulatory know-how to implement the expensive but necessary tests and make adequate restrictions. It could have stood as a model, showing its less prosperous Asian neighbours how to nip a potential epidemic in the bud. Instead, Japan has fallen behind. While other countries implement restrictions on feeding ruminants to ruminants, Japan has been content with non-binding 'administrative guidelines' which have reportedly been ignored in certain cases. Only after finding its first mad cow did Japan tighten up these guidelines into a compulsory regulation.

Japan criticized the European Commission's report because of an alleged lack of evidence, saying it would only alarm the public. "Lack of scientific evidence" was also the excuse made for not taking action in the earlier cases. The more likely reason is the tight connection between government and industry or government and healthcare institutions. These cosy relationships, often enhanced by a ministry official's retiring to a position in a large company, are cause for alarm.

Japan needs unbiased opinions, such as those the European Commission tried to provide. Perhaps then it would cease making inappropriate judgements that turn out to be fatal for some citizens. ■

Reviewing cancer and immunology

Introducing two new Nature journals.

Colorectal cancers are the third most frequent cancer worldwide, with increasing incidence in the developing world. More positively, they provide a vivid example of how serendipity and clinical observation can lead to fundamental biomedical insights. Long-term use of nonsteroidal anti-inflammatory drugs was serendipitously found to reduce colorectal cancer incidence in large clinical trials. By taking these clinical results back into the lab, researchers identified an enzyme, cyclooxygenase 2 (COX-2), that is overproduced in colorectal cancers and is clearly involved in tumour development. Selective COX-2 inhibitors reduce the number of polyps in people with familial adenomatous polyposis, an inherited predisposition to colorectal cancer, so may be of significant clinical benefit in preventing colorectal cancer in these individuals.

This synergy of disciplines, not to mention the vast flow of information that cancer specialists have to battle with, highlights a prime motivation of one of two Nature review journals being launched on

1 October. *Nature Reviews Cancer* aims to help all cancer researchers not only to develop an overview of their own areas of expertise but also to understand new opportunities emerging in others.

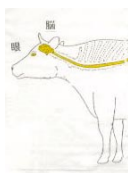
Interdisciplinary motivations have also stimulated the launch of *Nature Reviews Immunology*. Immunology is a diverse and growing discipline, but it is also at risk of fragmentation. It faces major challenges. The threat from autoimmune diseases is ever-increasing; we do not yet know how to stimulate the immune system to combat cancer; infectious diseases are on the rise and the need for efficacious vaccines is ever-present; and much remains to be learned about over-coming organ-specific transplant rejection. And the statistics speak for themselves: 30% of all deaths in 2000 were attributable to infectious diseases, and 2.7 million people alone died from HIV infection.

Immunologists and cancer researchers clearly have a lot of work to do. The new journals are intended to help them and their disciplines thrive. ■





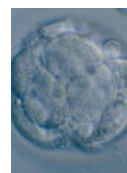
Facing the threat
US budget outlook
modified in wake
of terror attack
p336



Asia-bound
How Britain sent
mad cow disease
to the Far East
p337



X-ray vision
Beamlet laser helps
nuclear fusion on
its way
p338



Cell mates
Chinese scientists
put human nuclei
into rabbit eggs
p339

'Identity crisis' racks depleted research arms of spy agencies



William Triplett, Washington

The terrorist attacks of 11 September have thrown into sharp focus a set of problems that have been afflicting the research sections of US spy agencies for years, close observers say.

Agencies such as the Central Intelligence Agency (CIA) and the National Security Agency (NSA), which once led the world in their use of such technologies as satellite surveillance and supercomputing, have seen their technical edge dissipate since the end of the Cold War, according to the observers.

James Hirsch, who was deputy director of the CIA's directorate of science and technology from 1990 to 1995, says that intelligence-agency researchers "are in an identity crisis. They don't know who or what they



CIA director George Tenet and President Bush urgently need results from the troubled spy agency.

are or what they want to do because there's no clear sense of mission."

For half a century, spy technology was

geared towards the Soviet Union and its client states, entities that had vast military complexes and fixed lines of communication to spy on. The terrorists who conducted the 11 September attacks had neither of these. And despite signs during the past decade that terrorism was becoming more of a threat to US interests, US intelligence failed to identify and concentrate on terrorism — or anything else — as its principal target.

Instead, US spies and spycraft have simply become involved in several different problems in turn: crises in North Korea, Somalia, Haiti and the former Yugoslavia, for example. Lacking a clear mission focus, the agencies have been unsure what their research programmes are needed to prepare for, and these programmes have been shrinking as a result.

Although detailed figures are classified, the intelligence agencies are known to have an annual budget of approximately \$27 billion, of which an estimated \$3.3 billion goes to the CIA. Jeffrey Richelson, author of a recently published book, *The Wizards of Langley: Inside the CIA's Directorate of Science & Technology*, says that about two or three per cent of the agency's operating budget — \$80 million — was devoted to research and development. That's a low figure, by the standard of either corporations or government agencies with strong technical requirements. "There were plans to push it to 5%,"

Swift move to collect fallout data

Rex Dalton, San Diego

Environmental-health experts moved swiftly but sensitively to collect early exposure data on the possible fallout from the collapse of the World Trade Center in New York. They say such data will be vital in assessing any long-term health risks to tens of thousands of rescuers and survivors.

Immediately after the attack, officials at the National Institute of Environmental Health Sciences (NIEHS) moved to execute a research plan. Under NIEHS direction, researchers from at least four universities began collecting data on 12 September. They took samples from near the attack site, installed automated monitoring machines to collect air samples across New York City and beyond, and began preparations for long-term health surveillance.

Christopher Portier, director of toxicology at the NIEHS, says that after previous disasters, including fires, volcanic eruptions and industrial explosions, there has been too little early data collection.

"People were always a little too late getting started," says Portier, whose agency supports 30 environmental-health research centres at US universities.

Centres at Columbia University, Mount Sinai School of Medicine, New York University and Johns Hopkins University in Baltimore, Maryland, are participating. Researchers from the Environmental Protection Agency (EPA) and the Centers for Disease Control and Prevention were also collecting data after the attack.

Officials say that arsenic, mercury, lead, chromium, beryllium and titanium are among the problem substances to which rescuers and others may have been exposed. There was also some asbestos in one of the towers, although the EPA has said that the risk it poses is small.

With emergency personnel concentrating on saving lives, researchers faced immediate problems gaining access to the site, and at least one team decided to withdraw a week after the attack.

he says, "but it was one of those things that eventually requirements [elsewhere in the agency] ate up."

Last October the CIA, at least partially aware of the problem, overhauled the directorate and established a new Office of Advanced Technologies and Programs, according to agency documents. In 1999 it had set up In-Q-Tel, a non-profit organization designed to improve agency ties with start-up companies, where expertise in information technology is increasingly concentrated.

"The idea was to try and identify new, innovative information technologies that could be incubated with seed money from this organization," explains Hirsch.

In-Q-Tel was part of an effort by the CIA, which has no research laboratories of its own, to break its technical reliance on a small circle of contractors who know its way of working and conform with its secrecy requirements. The early results of the concept were "quite impressive", Richelson claims.

Far more rapid progress will now be expected, to get the CIA and other spy agencies abreast of terrorists' means of communication. Hirsch says that one typical requirement is for something similar to a web-browser, that would scour the US government's vast but fragmented and closely-held intelligence databases.

Security agencies have been tight-lipped about their ability to monitor the Internet, but outside experts say that blanket monitoring of all the traffic on it is out of the question. There are "just way too much data to handle," says Richard Clayton, a computer scientist with the Computer Security Group at Cambridge University.

The monitoring of traffic to and from certain e-mail addresses and geographical locations is more feasible, and is already undertaken by systems such as the FBI's Carnivore programme, which picks up messages by physically tapping Internet nodes and surveys them for keywords or phrases of interest.

One of the most powerful approaches to monitoring e-mail, however, is to locate one of the computers that sent or received it. Documents and even decrypted and deleted e-mail messages can then be retrieved fairly easily, says Jim Bates, managing director of UK-based Computer Forensics.

Bates says that even supposed computer experts would be unable to cover their cyber-tracks completely. He recalls one case when 98% of the e-mail messages relating to a banking fraud case were found on the computers that had sent or received them, even though the hard drives had been reformatted. ■

Additional reporting by David Adam

Support for science is firm as emergency raises budget



Matthew Davis, Washington

Faced with the terrorist attacks on New York and Washington, lawmakers have put aside disagreements over the dwindling budget surplus and agreed a bipartisan funding package that should increase spending on most science programmes.

The package would allow Congress to proceed with previously stalled funding bills for the 2002 fiscal year, starting on 1 October. If the deal is approved by the White House, it will provide, among other things, \$2.8 billion in further spending for the National Institutes of Health (NIH), and a boost of as much as \$400 million for the National Science Foundation.

The intensive effort to wrap up the bills next month is being assisted by a modest infusion of extra money. "The sense is that we have more money than we thought we would have just a few weeks ago," says one congressional aide, "but perhaps not as much as some ultimately would have hoped for at the end of the process."

Congress is already reviewing research and development programmes that could aid the US response to the terrorist attacks. It may choose to bolster them with a portion of the \$40 billion in emergency spending approved by Congress last week to deal directly with the attacks — or from further supplementary spending. Some have stated privately that the total bill for America's response to the attacks could come to several hundred billion dollars over an unspecified period of time.

President George W. Bush's administration was well on its way to finalizing the 2003 budget — which Bush will present next February — at the time of the attacks, but this will now be subject to drastic revision.

A House of Representatives panel that handles budgets for the NIH and the Centers for Disease Control and Prevention (CDC) met just three days after the attacks to review spending on programmes related to bioterrorism. The chair of the panel, Ralph Regula (Republican, Ohio), says that he expects Bush to allocate some of the emergency money to the CDC's civilian anti-bioterrorism activities, which involve both research and public-health initiatives. Regula thinks the Bush administration's plan to spend \$181 million on CDC measures to combat bioterrorism in 2002 may be inadequate.

It is unclear whether the NIH might obtain more funding for bioterrorism projects. The government was already set to



High security: bioterrorism work at the Centers for Disease Control is set to gain impetus.

increase its budget from \$49 million to \$93 million next year. The work, which includes sequencing the genomes of potential bioterror agents, is aimed largely at developing better vaccines and diagnostic tests.

According to government officials, other candidates for emergency funding include secret military and security-related research programmes ('black-box' programmes) at Department of Energy laboratories.

A congressional aide who works with science programmes notes that as Congress reviews research, there will be a strong temptation to portray programmes as being relevant to counter-terrorism. "Everyone will recast what they're doing so it looks indirectly or directly related to this," he says.

Before the attack, Mitch Daniels, director of the White House Office of Management and Budget, had been asking agencies to cut their budgets by 5% in 2003. But he had indicated that the White House would consider boosting non-biomedical research funding.

That will now have to be weighed against the requirements of a global military campaign, rebuilding in New York and Washington, and coping with the growing demands for social spending and revenue losses related to the ensuing economic downturn.

"The outlook for 2003 spending has been completely scrambled," said one of the congressional aides. "All bets are off." ■

Japan's first BSE case fuels fears elsewhere

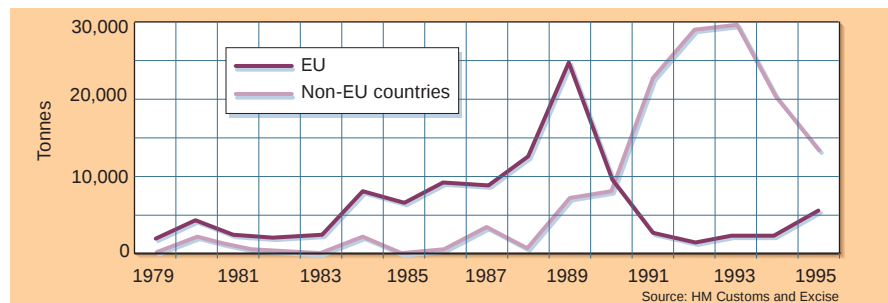
David Cyranoski, Tokyo

East Asian countries are bracing themselves for a possible wave of mad cow disease, after the first case of bovine spongiform encephalopathy (BSE) was confirmed in an animal born in Japan (see *Nature* **413**, 240; 2001).

The source of the infection is thought to be cattle feed containing animal parts, which British farmers dumped in the region when demand slumped in Europe in the early 1990s (see graph). Thailand and Indonesia were among the heaviest importers.

"If BSE got into cattle feed in Asia at that time, it's about time for the disease to show up," says Sumolya Kanchanapangka, associate professor of veterinary anatomy at Bangkok's Chulalongkorn University and adviser to Thailand's food-safety agency.

If the region has underestimated the risk of BSE, experts point out, its people could



UK exports of flours, meat pellets and meat offal, unfit for human consumption, during 1979-95.

be at risk of developing new variant Creutzfeldt-Jakob disease (vCJD), the neurodegenerative disease thought to be caused by eating brain, spinal cord and other organs from BSE-infected cattle. Workers in Thai slaughterhouses don't seem very thorough about removing these, says Kanchanapangka. In Korea, where "cow brain is a delicacy and every part of the cow is eaten, reports of BSE would lead to panic", says an agriculture ministry representative there.

Japanese government and industry officials previously argued that Japan was at little risk of BSE because the imported British meat and bone-meal was fed only to pigs and chicken, not to cattle. Japanese cattle are normally raised on beer and soy-based feed. But the confirmed case shows that this was not always so.

Earlier this year, when reports circulated that a European Commission study group was poised to report that Japan was at high risk of the disease, Japan stopped

cooperating with the study, effectively blocking it, according to a European official.

Other Asian animal-health officials still say that the animal feed was not given to cattle. Thai and Korean officials say that only pig and chicken farmers could afford the imported feed. But the European official, who declined to be identified, said that experience in Europe suggests that such restrictions don't always work.

Korean and Thai officials point out that, unlike Japan, their countries have no documented cases of scrapie, a disease of goats and sheep that is thought to be related to BSE. And the procedure of reprocessing animal parts for feed is rare in the region, they say.

Korea, Singapore and other neighbouring countries have responded to the Japanese case with immediate bans on the import of Japanese beef. Japan, meanwhile, says it will run BSE tests on one million cows, at a cost of some ¥3 billion (US\$25 million).



Grave disclosure: Japan's agriculture minister, Tsutomu Takebe, breaks the news.

Transgenic corn found growing in Mexico

Rex Dalton, San Diego

Genetically modified corn (maize) has been found growing in Mexico, raising sensitive environmental and cultural issues in the part of the world where the crop was first cultivated centuries ago.

Transgenic corn is widely sold for consumption in Mexico, where more than five million tonnes of corn are imported annually from the United States. But none of the corn is grown commercially there following a 1998 government moratorium.

The disclosure of scattered plots of transgenic corn in the states of Oaxaca and Puebla was made by a government official earlier this month. A research team at the University of California at Berkeley, which is preparing work on the topic for publication, has subsequently accused the official of breaching confidentiality by his disclosure. Preliminary results from a government study appear to confirm the transgenic corn.

Oaxaca, a rural southern state where maize is revered by indigenous people, is the global centre of corn diversity, and the place of origin of strains grown commercially around the world. Environmentalists claim that the arrival of transgenic strains there could disrupt the genome of naturally bred corn.

Reports released in Mexico City last week say that the existence of growing genetically modified corn was discovered by a team led by Ignacio Chapela, a plant molecular biologist from the University of California at Berkeley who has long worked in Oaxaca. A native of Mexico, Chapela confidentially shared preliminary results of his research earlier this year with Mexican government officials. The officials then set up a research team to conduct similar studies.

On 4 September, at a subcommittee meeting of an international food-safety

organization, the Codex Alimentarius Commission, Chapela's discovery was revealed publicly by Fernando Ortiz Monasterio, director of Mexico's biosafety commission. Within days, the information had reached the Mexican Congress and the press.

Chapela says he had told Ortiz and other Mexican officials that he was planning to publish his research, and that public disclosure would undermine this. He adds that Ortiz's "breach of confidentiality" will "degrade the quality of information" his team was compiling.

Ortiz denies breaching confidentiality, but acknowledges that he did reveal Chapela's research results in a public forum.

On 17 September, the Mexican environment ministry released partial results of its own study, which revealed that transgenic corn was found in 15 of 22 areas tested in Oaxaca and Puebla.

British archaeology boost may uncover old enthusiasm

David Adam, London

Archaeologists are hoping that a £1.2-million grant from the Leverhulme Trust will help to revive the study of ancient settlement patterns in the British Isles.

The grant will support five years of study by researchers at the Natural History Museum, London, and partner institutions. The team will investigate when people first arrived in Britain, and how the populations responded to environmental change. A particularly interesting question is why the country seems to have been uninhabited between 170,000 and 70,000 years ago.

Britain has some of the world's best archaeological sites from the Palaeolithic period, from about 500,000 to 12,000 years ago. In 1993, for example, a 500,000-year-old human shin bone was found at Boxgrove, near Chichester in West Sussex (see *Nature* 369, 311–313; 1994). But funding has been scarce for archaeology in recent years, researchers complain, and gravel extraction and coastal erosion are now threatening some of the most promising locations for exploration.

Archaeology “really has been neglected in the United Kingdom and many sites are being destroyed”, says Simon Parfitt of the Institute of Archaeology at University College London. He hopes the grant will help to revive interest in the subject.

Chris Stringer, head of human origins at the Natural History Museum and the project's director, says that Britain's variable geography and climate make it ideal for studying how early humans adapted to environmental change. Stringer says the project will conduct some new excavations, but will focus mainly on re-examining the extensive collections of stone tools, fossil remains and animal bones that already exist at various universities and museums. ■



Bone ideal: Chris Stringer says Britain is a good place to study early human settlements.

Crushing victory could help in quest for fusion energy



Laser gazer: part of the Beamlet, which is used to capture images of material compressed by X-rays.

William Triplett, Washington

Fusion researchers at Sandia National Laboratory in New Mexico have shown that an X-ray machine, known as the Z-machine, can be used to compress a small plastic pellet uniformly, bringing hopes of achieving controlled nuclear fusion a step closer.

The researchers want to develop the Z-pinch, as the machine's technique is known, to produce fusion energy. The technique already generates the world's most powerful laboratory X-rays. But until last month, the researchers hadn't been able to observe directly that the X-ray flash produced by their machine crushes its target pellet evenly.

They have now confirmed this by reconstructing the powerful Beamlet laser and using it to obtain images of the compressed pellet. Beamlet was first built at the Lawrence Livermore National Laboratory in California as a prototype for the National Ignition Facility now under construction there.

The laser image shows that the pellet was compressed uniformly, by a factor of two. “Pellets have been compressed before by more than that, and they will have to be compressed a lot more to achieve nuclear fusion,” says Stephen Dean, president of Fusion Power Associates, a Maryland-based group that promotes fusion energy. But he says the finding is “a real advance” for the New Mexico laboratory.

The Sandia researchers now want to increase the laser energy during compression so that the temperatures and pressures exerted on the deuterium-filled pellet are closer to those at which fusion will occur.

Seen from above, Sandia's Z-machine looks like a wagon wheel, some 40 metres

across. Its rim is surrounded by huge power capacitors, and the spokes consist of 36 aluminium and steel conductors that take power to the centre. There it is fed into extremely fine tungsten wires strung along the sides of a small cylinder.

When fired up, the capacitors send a current of 20 million amps through the conduits and into the tungsten wire. In the instant before the wires evaporate, the electrical current produces an extraordinarily powerful magnetic field, which in turn generates X-rays. The project's researchers think these X-rays could be channelled to produce controlled fusion, just as X-rays produce fusion in a hydrogen bomb.

Controlled nuclear fusion would require compression by a factor of about 20, says John Porter, manager of the Sandia project. He says he is confident that the Z-machine can achieve a factor of at least 10.

Sandia researchers spent \$13 million reconstructing the Beamlet for the Z-machine. Having used it to photograph the pellet, they want to shorten its laser pulse by a factor of 1,000 and use it to inject extra heat into the pellet. This project would cost a further \$30 million, the researchers say.

But even with better compression and extra heat, the existing equipment would probably fall short of achieving the self-sustaining fusion reaction that fusion researchers call ignition. “Theoretically [ignition] is within the realms of possibility,” says Porter. “But we'd be doing this more as a proof of concept than an attempt to start a real fusion burn. We are excited by the possibilities, yes, but we do have to temper that with reality.” ■

China plans 'hybrid' embryonic stem cells

Alison Abbott & David Cyranoski

The controversy surrounding research on human embryonic stem (ES) cells now extends to the world's most populous nation.

Chinese scientists have revealed that they have transferred nuclei from human cells into rabbit eggs stripped of their own chromosomes. They hope to use the resulting embryos to generate ES cells for research into regenerative medicine. Previously, only one other group — at the biotech company Advanced Cell Technology (ACT) in Worcester, Massachusetts — had admitted to transferring human cells into the eggs of another species in this way.

ES cells may allow the production of specialized cells and tissues to replace those lost to disease — neurons to treat sufferers of Parkinson's disease, for example, or cardiac muscle for heart-attack victims. If the embryos that provide these cells are created by the nuclear-transfer cloning technique used to create Dolly the sheep, the tissues would be genetically identical to the patient, avoiding the problem of immune rejection.

But this 'therapeutic cloning' requires eggs, and human ones are in extremely short supply. Three years ago, ACT announced that it had devised a way around the problem — replacing human eggs with those of cows, which can be obtained from cattle ovaries collected in abattoirs. The announcement provoked intense controversy, however, with many commentators expressing distaste at the idea of cross-species nuclear transfer.

This month, Chen Xigu of Sun Yat-Sen University of Medical Sciences in Guangzhou told newspapers that his research group is working along the same lines. Chen and colleagues removed the chromosomes from rabbit eggs and replaced them with the nuclei of skin cells from a seven-year-old boy. In some of the 100 or so successful nuclear transfers, an embryo developed to the morula stage — the compact ball of cells that forms after about three days of embryonic development.

Before ES cells can be isolated from human embryos, they need to develop beyond this stage to form a hollow, fluid-filled structure called a blastocyst. Chen has not yet been able to coax his embryos into developing this far, but he is optimistic of success.

However, Jose Cibelli, who leads the ACT therapeutic cloning team, says that growing the blastocysts will not be easy. "There is a big difference between getting a nice morula stage and the next step," he says. Cibelli's team has managed to create blastocysts, but only at a low rate. They take ten days or so to develop, compared to five or six days in normal human embryos.

Chen plans to submit his work for publi-



Key point: human embryos must develop beyond the morula stage before stem cells can be isolated.

cation by the end of this year, but the ACT team does not intend to publish. "We are not happy with our results, which are inconsistent, and we are not happy with the general quality of the blastocysts," says Cibelli.

Chen's work has attracted criticism in China. "Some people in China are against me doing this research," he admits. And pressure is increasing to introduce legislation.

Zhu Chen, vice-president of the Chinese Academy of Sciences and director of the publicly-funded Chinese National Human Genome Center in Shanghai, has been campaigning for tighter ethical regulation of biomedical research. After reading newspaper reports about Chen Xigu's work two weeks ago, he called on the National Centre for Biotechnology Development, part of the Chinese Ministry of Science and Technology, to propose ethical guidelines for genomic and stem-cell research. "It is of great importance for a large country like China for this kind of work to be under strict regulation," he says.

Ming-Wei Wang, a veteran of Shanghai's biotech industry and vice-chairman of Ancile Pharmaceuticals in San Diego, says the regulation of biomedical research in China is patchy, even where laws are in place. ■

Stamp booklet has physicists licked

Erica Klarreich, London

British physicists are taking issue with a Royal Mail brochure, published in association with six new postage stamps, which suggests that quantum physics could help to explain the paranormal.

The brochure, part of a presentation package to commemorate the centenary of the Nobel prizes, will be available at 18,000 post offices from 2 October.

The material at issue was written by Brian Josephson, a professor at the University of Cambridge, who a Nobel prize in 1973 for his work on superconductors. The entry says that developments in quantum physics, combined with information technology, "may lead to an explanation of processes still not understood within conventional science, such as telepathy — an area in which Britain is at the forefront of research".

Josephson gave his name to the Josephson effect, which describes the flow of current between two pieces of superconducting material separated by a thin layer of insulating material.

But few physicists accept that telepathy even exists, says Andrew Steane, a quantum physicist at the University of Oxford.

Robert Evans, a



physicist at the University of Bristol, says he is "very uneasy" about something from the Royal Mail saying quantum physics has something to do with telepathy," he says.

Kathryn Hollingsworth, a spokeswoman for the Royal Mail, says: "If it transpires that what he's suggesting doesn't have a scientific basis, perhaps we should have checked that," she says. "But if he has won a Nobel prize for his work, that should give him some credibility."

Josephson wrote the paragraph at the Royal Mail's invitation. A number of theories, he says, are helping to explain "how a more detailed understanding of paranormal phenomena may emerge from a better understanding of quantum mechanics." ■



Small-scale science scoops large grants for six new centres

Washington The National Science Foundation has awarded grants worth US\$65 million over five years to support six major new university-based centres for nanoscale science.

The centres will investigate nanoscience and its applications in fields that include electronics, materials, medicine and environmental engineering. They will be located at Harvard University in Massachusetts, Northwestern University in Illinois, Rice University in Texas, and three institutions in New York state — Columbia and Cornell universities and the Rensselaer Polytechnic Institute.

“Together they will provide coherence and a longer-term outlook to US nanotechnology research and education,” says Mihail Roco of the agency’s nanotechnology office.

Australia moves forward on stem-cell research

Sydney Australia took a step towards harmonizing its rules on human embryonic stem-cell research last week when a Federal Parliament committee recommended that

such research be allowed in all states.

But the committee was split over how the cells should be obtained. There was unanimous support for research on existing stem-cell lines. Australian company BresaGen, based in Thebarton, South Australia, is one of the few in the world to hold such lines. But only six out of the ten committee members supported the extraction of stem cells from embryos left over from *in vitro* fertilization procedures.

The committee also called for a ban on the deliberate creation of embryos for experimentation or reproductive purposes. A three-year moratorium on somatic-cell nuclear transfer was suggested, during which the scientific benefits would be assessed. The report has yet to receive an official response from the government, but seems likely to be implemented.

Biomedicine to gain from institute merger

Singapore The rivalry between two of Singapore’s top life-sciences institutes is set to end after a high-level government advisory committee gave its blessing to a proposal for their merger.

The Institute of Molecular and Cell Biology and the Institute of Molecular Agrobiology (IMA) have in the past been



Under pressure: Singapore’s agrobiological institute has been subjected to criticism.

bitter rivals (see *Nature* **412**, 370–371; 2001). But the IMA has recently come under pressure, with critics arguing that a country with little agriculture should not be investing in such research. Rising opposition to genetically modified crops in East Asia has further weakened the IMA.

Merging the two institutes, which employ approximately 200 scientists each, will create “a critical mass” for research in molecular biology and biomedicine, according to Philip Yeo, co-chairman of Singapore’s Economic Development Board and a main architect of the country’s efforts to nurture an indigenous biotechnology industry. The merger is due to be completed by 2003.

Drinking-water arsenic update shows increased risk

Washington Cancer risks associated with arsenic in drinking water may be higher than previously thought, according to a new study by the US National Academy of Sciences.

An upper limit for arsenic levels of 10 parts per billion (p.p.b.), down from the present 50 p.p.b., was set during the last week of the Clinton administration. But the Environmental Protection Agency (EPA), which is responsible for regulating water quality, scrapped the plans this March, citing insufficient scientific evidence for the change and saying that an independent scientific review was needed (see *Nature* 410, 503; 2001). The academy has now found that the risk of cancer is 10 to 15 times greater than the EPA considers acceptable, even at levels below 10 p.p.b.

The new analysis, *Arsenic in Drinking Water: 2001 Update*, is an update to a 1999 report, and uses new information, a different study design and more sophisticated computer models to quantify the risks at low levels where effects had not previously been discernible. Evidence from the study will be used by the EPA when it decides on a new standard next month.

► <http://www.nap.edu>

Malaria drug set for use in CJD trials

London Clinical trials of a headline-making treatment for Creutzfeldt–Jakob disease (CJD) are to go ahead in Britain, after the government asked the Medical Research Council to test the efficacy of the antimalarial agent quinacrine against the condition.

The move follows reports in the UK press last month that researchers at the University of California, San Francisco, had used quinacrine to successfully reduce the symptoms of a British woman suspected of having the disease.

“The decision to evaluate quinacrine does not mean that we regard it as a ‘front runner’ in finding an effective treatment for CJD,” says Liam Donaldson, chief medical officer for the Department of Health. “It simply means that any research claim that might offer hope in an otherwise fatal disease must be evaluated.”

Snakebite kills herpetology researcher on jungle trip

San Diego A herpetologist died last week on an expedition to the remote jungle of Myanmar (formerly Burma) after he was bitten by a deadly snake. Joseph Slowinski



Slowinski: bitten by a snake being removed from a capture bag.

during his 11 expeditions to Myanmar.

of the California Academy of Sciences in San Francisco was conducting research paid for by the National Science Foundation when a small krait snake struck while being removed from a capture bag.

Slowinski is known for identifying at least 18 new species of reptiles and amphibians

Correction Last week's News story on counter-terrorist technology (see *Nature* 413, 238; 2001) incorrectly described PerkinElmer Instruments' Threat Image Projection system as a software package that relies on pattern-recognition algorithms to spot weapons and other suspicious objects during X-ray baggage screening. In fact, the system is designed to monitor and improve the watchfulness of airport staff by randomly projecting images of such objects onto X-ray images of screened baggage.

All creatures great and small

The living world is governed by laws based on fractal geometry and on the sizes of organisms, some scientists claim. John Whitfield looks at the debate surrounding a biological 'theory of everything'.

Some people see only the trees, others the whole wood. But when Brian Enquist enters one of his Costa Rican field sites, he senses far deeper patterns. "When I walk into the forest, I have the feeling that, although it's very complex, there are simple rules underlying that complexity," he says.

Enquist, an ecologist at the University of Arizona in Tucson, is working to derive laws that can explain the workings of ecosystems, and the biology of their constituent organisms, in terms of those organisms' sizes — a project "as potentially important to biology as Newton's contributions are to physics", as one commentator recently put it¹. By explaining simple scaling laws — mathematical expressions of how organisms' biology varies with their size — in terms of the fractal geometry of networks such as circulatory systems, Enquist and his colleagues hope to understand patterns in metabolism, growth, ecology and evolution right across the living world.

A string of highly cited papers in top journals, and the Ecological Society of America's

2001 Mercer Award for young investigators, to Enquist, attest to the impact of these ideas. But, as with any bold new theory, some question its generality, and others its truth. Other critics even deny the existence of the very trends that the theory seeks to explain. Have Enquist and his colleagues really made a stunning theoretical breakthrough, or is the comparison with Newton hype? The debate rages, but whatever consensus eventually emerges, just about everyone involved believes that important biological insights will stem from this new focus on scaling.

Scale model

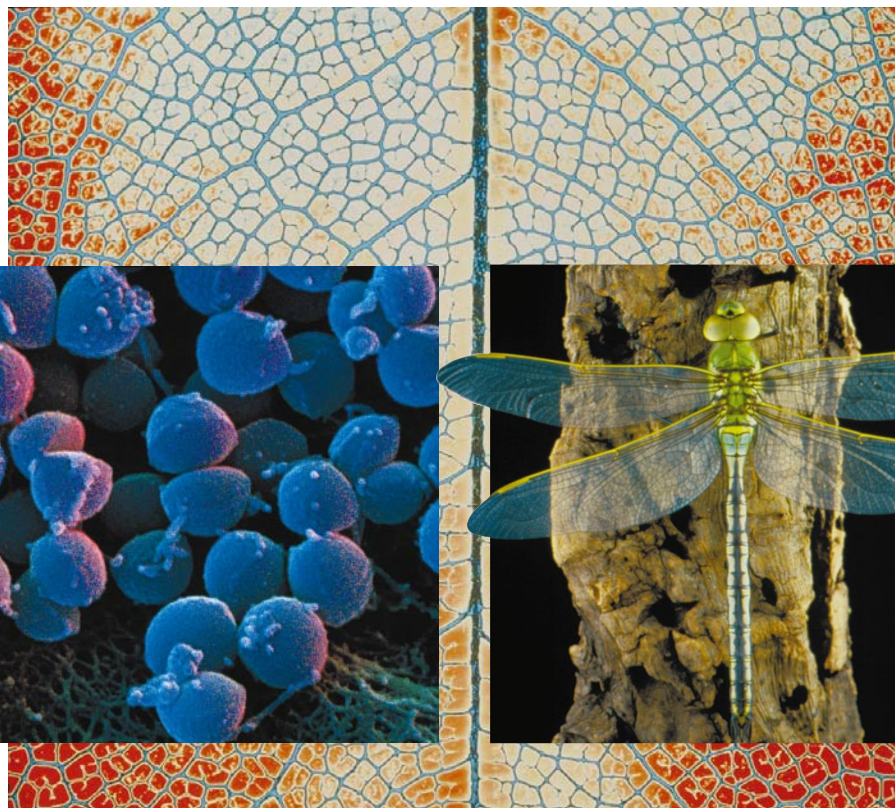
In attempting to derive a fundamental explanation for biological scaling laws, Enquist and his two associates — ecologist James Brown of the University of New Mexico in Albuquerque, and physicist Geoffrey West of Los Alamos National Laboratory, also in New Mexico — have reopened the files on a mystery that has baffled biologists for more than half a century.

Naturalists have long known that many

aspects of organisms' biology vary with their size. Bigger animals live at a slower pace. They survive longer, grow more slowly, have slower heart rates, and so on. In the 1930s, physiologist Max Kleiber of the University of California, Davis, put a number on this trend. He showed that an animal's metabolic rate is proportional to its body mass raised to the power of $3/4$ (ref. 2). This relationship has been found to hold across the living world from bacteria to blue whales and giant redwoods, over more than 20 orders of magnitude difference in size. Scaling laws based on exponents in which the denominator is a multiple of four apply to a host of other biological variables, such as lifespan. "There's maybe 200 scaling laws that have quarter powers in them," says West.

But for decades, scaling laws stubbornly resisted explanation. Organisms are three-dimensional. So, from the principles of euclidean geometry, one might expect that biological scaling laws would operate on multiples of the third, rather than the quarter power. Metabolic rates, for instance, might scale as the $2/3$ power of body mass — this is how body surface area, where metabolic heat is lost, scales with its volume, which determines how much metabolic heat a given organism can produce.

In the mid-1990s, Enquist and Brown began to search for factors that might explain



the 'missing' fourth dimension of biological scaling. They suspected that the dynamics of organisms' internal transport of nutrients and other resources might provide the key. Through contacts made at the Santa Fe Institute in New Mexico, which specializes in thorny cross-disciplinary problems, they teamed up with West, a theoretical physicist with the mathematical expertise to help them develop the idea.

The resulting theory frames the scaling conundrum in terms of resource-distribution networks, such as blood vessels, the xylem that transports water through plants and the tracheal tubes that carry oxygen to an insect's tissues. Thus, variables such as metabolic rate — and the effect on them of changing body size — become a consequence of how resources are shunted around these networks.

West, Brown and Enquist's theory starts with the assumption that evolution has made biological resource-distribution networks maximize the area across which they can take up and release resources and minimize the time and energy needed to transport those resources through the organism. "Natural selection is very powerful," says Brown. "It's difficult for organisms to deviate from the optimum." The model also assumes that the size of networks' terminal units is independent of body size — that a mouse's capillaries, say, are the same size as those of an elephant.

Networks that fulfil these criteria turn out to have fractal geometry, argue West, Brown and Enquist. This means that their branching structure can be described mathematically by applying a simple mathematical formula over and over again. Abandoning Euclid and embracing fractals was the key step, says Brown. "Much of life is designed in a fractal-like way. We were able to make that cogent and

relevant by developing rigorous quantitative models."

This fractality seems to solve the enduring puzzle of why the scaling exponents of three-dimensional organisms are multiples of the number four. Filling a three dimensional volume with a network that maximizes the two-dimensional surface area available for capturing and releasing resources creates a four-dimensional geometric entity³. This is difficult to visualize, but Enquist uses the analogy of how a tree uses three-dimensional space to cram a much larger area of essentially two-dimensional leaves into the ground space covered by its canopy. "Fractal geometry kicks life up to another dimension," he concludes. In April 1997, West, Brown and Enquist unveiled their theory by showing that they could derive Kleiber's 3/4-power scaling law linking body mass and metabolic rate⁴.

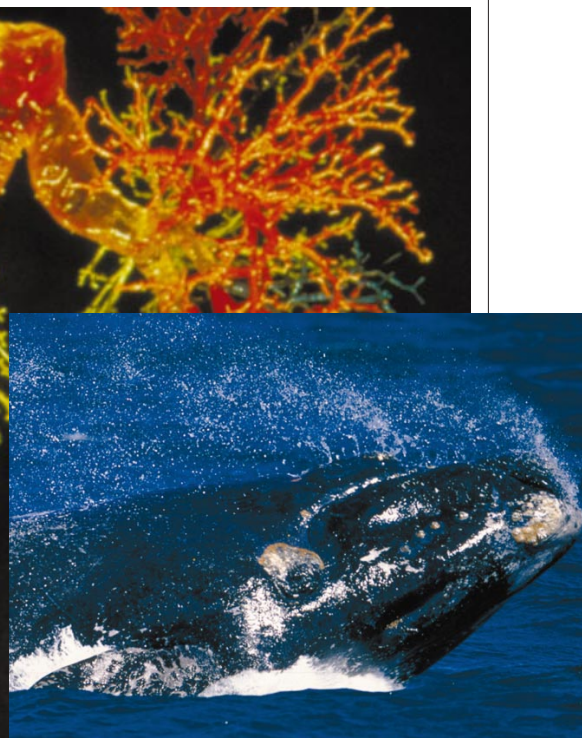
Under fire

The researchers have since used the theory to describe a range of biological phenomena, such as scaling and structure in vascular plants⁵. They have also expanded their thinking from the level of individual organisms to entire ecosystems. Using their models describing the relationship between body mass and metabolic rate, for instance, the researchers have shown that — wherever resources such as nutrients or water are a limiting factor — the population density of trees scales to the $-3/4$ power of each individual's mass⁶. The rule seems to hold for forests from the Amazon to the Arctic⁷.

The same theoretical framework also explains why, regardless of the speed at which plants' diameter or height increases, their growth rate scales to the $3/4$ power of their body mass⁸. According to this view, different plant life histories, encompassing vastly different rates of growth and timings of sexual maturity, simply represent different ways of following the same law for optimal energy use⁹.

So have West, Brown and Enquist derived biology's 'theory of everything'? Enquist sees their current ideas as "an excellent first approximation" of an eventual general explanation of resource use by organisms. But the theory has come under fire from some researchers who question its initial assumptions, and from others who believe that biological reality has been lost in the quest for generality.

Earlier this year, a group of physicists at the Massachusetts Institute of Technology questioned the very existence of the $3/4$ law for mass and metabolic rate. Re-analysing published data sets dating back to Kleiber's original, Peter Dodds and his colleagues concluded that the exponents in most of them were statistically indistinguishable from $2/3$ (ref. 10). "The original data on metabolism don't really bear out the prevailing idea that it's a $3/4$ exponent — you get all sorts of scalings," claims Dodds, who now works at New York's Columbia University. Faced with a cloud of decimal places from different studies, earlier generations of biologists fudged on $3/4$, partly because it made their slide-rule calculations easier, Dodds suggests.



Brown denies that the numbers aren't up to scratch. "Metabolic rate is not difficult to measure," he says. "And whenever we have a reasonable number of measurements, they all cluster about $3/4$." William Calder, an ecologist at the University of Arizona, says that the large number of biological variables with $3/4$ -power scaling confirms the trend. "It all hangs together," he says.

Biologists working on fine details of the biology of particular organisms have criticized the fractal theory's attempts at generality. Although most agree that the cross-species correlations look good overall, many individual species deviate widely from the scaling rules. "The more detail that one knows about the particular physiology involved, the less plausible these explanations become," says Henry Horn of Princeton University in New Jersey, a botanist who studies patterns of tree growth. "I have a tendency to think that nature does different things in different ways."

But criticism based on the observation that individual species disobey scaling laws misses the point, says West. The theory is designed to capture laws true for an idealized average organism, rather than every individual plant and animal. West says that in fact he expected to find that many more species deviated from the theory's predictions. "What I find most mysterious is why in hell does it work so well? It's a little bit frightening."

West puts much of the disagreement down to a cultural difference between physicists and biologists. "If Galileo were a biologist, he would have written a big fat tome on the details of how different objects fall at different rates," he says. Enquist agrees. "Biologists tend to lack a philosophical bent to look for generalities," he says.

Branching out

But growing numbers of biologists are now gravitating towards the new theoretical framework. "I've gone from being an atheist to an agnostic," says botanist Karl Niklas of Cornell University in Ithaca, New York. Niklas, the author of the newtonian comparison, now collaborates with Enquist. If the theory holds up, he says, it has "the potential to explain virtually everything" that relates to organisms' sizes.

Another enthusiast, ecologist Mark Ritchie of Utah State University in Logan, has adapted the theory to explain how animals of different sizes share out the resources in an ecosystem. He has modelled communities of foraging animals as a fractal network of resource exploitation. From this, he has predicted how many species can coexist in a particular place — in other words, he is trying to explain biodiversity from fundamental principles. The model gives a good fit to the patterns of species richness seen on the Serengeti and the Minnesota prairie¹¹. In unpublished work, Ritchie has moved on to



Seeing the wood and the trees: Brian Enquist is searching for a grand unity in nature.

looking at how the area of an animal's home range scales with body mass. For carnivores, he has found that the amount of land needed rises sharply with increasing body size, because their prey is relatively rare. This makes predators especially vulnerable to habitat fragmentation.

Even critics such as Horn do not argue with the data that West, Brown and Enquist are producing. "The relationships are so tight that there have to be some very powerful generalizations behind the mechanics," he says. But Horn doubts the significance of fractal geometry. Although trees look like fractals, he says, their growth is more complicated than the iteration of mathematical formulas, involving a complex balance between tissue formation and death.

Jayanth Banavar, a theoretical physicist at Pennsylvania State University in University Park, believes biological generalizations can be explained without resorting to fractals. Two years ago, Banavar and his colleagues published a simpler model, which argued that Kleiber's $3/4$ rule is a consequence of any network in which resources flow outward from a source to their uptake sites¹². Considering the total length of the network, the number of sites at which resources can leave it, and the total amount of resources in the network at any one time, they argued that the $3/4$ rule falls naturally from the calculations.

"The three-quarters is inherently built into any directed network — you don't need fractality, or optimality, or anything," says Banavar.

But many experts on scaling are sceptical of Banavar's approach — including those who are critical of West, Brown and Enquist. West adds that Banavar's theory does not explain the wide variety of scaling phenomena embraced by the fractal theory. "It's not correct," he claims. "And even if it was correct it only explains one piece — the metabolic rate."

Meanwhile, like a fractal, West, Brown and Enquist's theory is branching out in all directions. "We were aware that we'd uncovered the tip of an iceberg," says Brown, "but not how big it was and where it might lead."

Brown, Enquist, West and their collaborators now have their sights set on a mind-boggling range of problems. They believe that scaling laws can be applied to the molecules and subcellular organelles involved in respiration and cell division. A paper co-authored by Brown and published last week includes trends in body temperature — the next most important biological property after size, says Brown — in the theory¹³. Eventually, he hopes to explain relationships between temperature, body size and rates of evolutionary change.

Brown is also intrigued that animal cells grown in a culture dish increase their metabolic rate and rate of division — as if they were aware of no longer being part of a massive unit. The same goes for cancer cells, which have rebelled against the larger unit. He is similarly fascinated by the observation that cancer cells can modify resource-distribution networks, stimulating blood vessel growth.

Whether West, Brown and Enquist have hit the nail on the head with their fractal approach, Niklas is convinced that simple and general rules governing diverse biological phenomena lie waiting to be discovered. "I don't think that the ecological patterns that we see surfacing in fossils and living organisms and across the continents are a consequence of chance," he concludes. ■

John Whitfield works in Nature's science writing team.

1. Niklas, K. J. *Trends Ecol. Evol.* **16**, 468 (2001).
2. Kleiber, M. *Hilgardia* **6**, 315–353 (1932).
3. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **284**, 1677–1679 (1999).
4. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **276**, 122–126 (1997).
5. West, G. B., Brown, J. H. & Enquist, B. J. *Nature* **400**, 664–667 (1999).
6. Enquist, B. J., Brown, J. H. & West, G. B. *Nature* **395**, 163–165 (1998).
7. Enquist, B. J. & Niklas, K. J. *Nature* **410**, 655–660 (2001).
8. Niklas, K. J. & Enquist, B. J. *Proc. Natl Acad. Sci. USA* **98**, 2922–2927 (2001).
9. Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. *Nature* **401**, 907–911 (1999).
10. Dodds, P. S., Rothman, D. H. & Weitz, J. S. *J. Theor. Biol.* **209**, 9–27 (2001).
11. Ritchie, M. E. & Olff, H. *Nature* **400**, 557–560 (1999).
12. Banavar, J. R., Maritan, A. & Rinaldo, A. *Nature* **399**, 130–132 (1999).
13. Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. *Science* **293**, 2248–2251 (2001).

SPL

Storm in a culture dish

Will restrictions on embryonic stem-cell research and attempts to ban therapeutic cloning erode the United States' leadership in biomedical research? Laura Bonetta considers the evidence.

"Driven by ignorance, conservative thinking and fear of the unknown, our political leaders have undertaken to make laws that suppress this type of research ... I believe our country risks being thrown into a dark age of medical research."

With these words, delivered in an opinion article in *The New York Times* on 31 August, software mogul Jim Clark announced that he was suspending \$60 million of a \$150-million donation to Stanford University in California (see *Nature* 413, 5; 2001). He attacked

restrictions imposed on federal funding for research on human embryonic stem (ES) cells, and legislation before Congress that would ban human therapeutic cloning — a technique that could generate ES cells genetically matched to individual patients.

These arguments echoed statements made in July by stem-cell researcher Roger Pedersen, when he announced that he was leaving the University of California, San Francisco, for the University of Cambridge (see *Nature* 412, 262; 2001), citing Britain's political support for ES-cell research.

Meanwhile, Geron of Menlo Park, California, the leading company in the field, began telling reporters that the bill before Congress opposing therapeutic cloning might force it to shift investment to its British-based subsidiary Geron BioMed. In its 10 September issue, *Business Week* summed up the worries about US competitiveness with an article under the headline: "At risk: a golden opportunity in biotech".

Despite the gloomy headlines, most experts interviewed by *Nature* do not believe that there is an imminent danger of a brain drain of top ES-cell scientists from the United States to countries that have taken a more permissive attitude to the research, nor of an immediate threat to US competitiveness in biomedicine more generally.

First, they stress the need to put the threats into perspective. Clark's money, for instance, is being used to construct an interdisciplinary research centre called Bio-X, but its construction had already been scaled back in the face of opposition from local residents. And the generous salaries and enviable working conditions available for

Under the lens: the US Congress may ban therapeutic cloning.

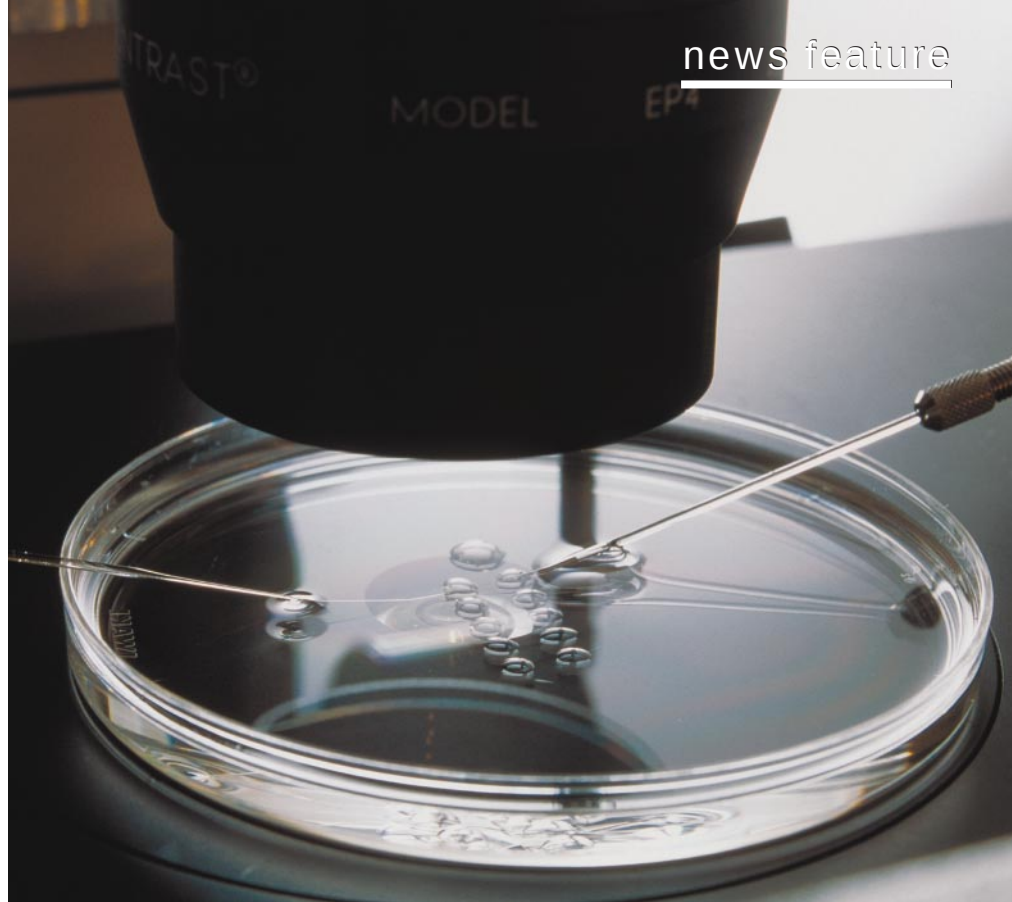
US biomedical scientists mean that a significant brain drain is unlikely.

Making a start

More importantly, federally funded US researchers can now at least make a start on human ES-cell research, thanks to the compromise, announced by President George W. Bush on 9 August, that allows researchers to seek grants from the National Institutes of Health (NIH) to work on 64 human ES-cell lines isolated before then (see *Nature* 412, 665; 2001). Work on human ES cells is at such an early stage that US researchers will for now have plenty to do characterizing these lines and trying to determine whether any of them have clinical potential.

"I don't think there will be a large-scale migration of top US researchers to the UK, Singapore, Sweden and Australia," says Alan Trounson of the Institute of Reproduction and Development at Monash University near Melbourne, Australia, where six of the these cell lines were derived. The main concern, however, is whether controls on federal funding will restrict the number of US researchers entering the ES-cell field.

Much will depend upon how the entire field of regenerative medicine develops. From 1988 to 1993, the administrations of Presidents Ronald Reagan and George Bush Sr imposed a moratorium on federal funding for research on therapies involving tissues from aborted fetuses. But because



Forbidden fruit: will US researchers be denied access to the best cell lines?

► trials using fetal tissue grafts to treat conditions such as Parkinson's disease have yielded disappointing results, that ban is, in retrospect, not felt to have kept US researchers out of a crucial area of biomedicine.

Most experts believe that stem-cell therapies hold much greater promise — in theory, it should be possible to use them to repair a wide range of the body's tissues. But, at present, it is unclear whether ES cells will emerge as the main clinical contenders, or whether they will be eclipsed by therapies that exploit the adult stem cells found in many of our tissues. The prospects for therapeutic cloning, which may be too expensive and cumbersome for routine use, are even more uncertain (see *Nature* 410, 622-625; 2001).

Attention will now focus on practical questions surrounding the approved human ES-cell lines. Experience with mouse ES cells suggests that labs will, over time, zero in on a handful of lines that remain stable and viable, and retain their ability to differentiate into each of the body's tissues. If those are all on the list approved by President Bush, US researchers may be able to compete effectively in the initial stages of basic research.

"I think most will watch closely and give the situation another year to mature," says Tom Cech, president of the Howard Hughes Medical Institute (HHMI), based in Chevy Chase, Maryland, the largest US source of philanthropic funding for biomedical research.

But there is a strong possibility that new cell lines may emerge that are clearly superior to those currently approved for US federal funding. "At this stage, we cannot predict where discoveries will come from," says Chris Juttner, vice-president for clinical affairs with BresaGen of Adelaide, Australia, which developed four of the ES-cell lines on Bush's list. He notes that several companies are trying to establish human ES cells that

can grow in the absence of a layer of mouse 'feeder' cells, making them more suitable for use in human patients. "It is only a matter of time that scientists figure out a better way to derive stem cells," says Lawrence Soler, chair of the Coalition for the Advancement of Medical Research in Washington, which lobbies in favour of ES-cell research. "As time goes on, there will be more and more frustration among American scientists."

Twin track

US academics can work on other human ES-cell lines if they obtain private funds. Harvard University and the HHMI recently announced that they will jointly fund a team to isolate further lines of human ES cells. But such arrangements create severe administrative and logistical problems — the Harvard scientists will have to ensure that no federal money is used to pay for any reagents or overhead costs.

Geron has supported the work of academic groups for several years, including the team led by James Thomson of the University of Wisconsin in Madison who isolated the first human ES-cell lines three years ago. To deal with the problem of keeping Geron-sponsored research separate from federally funded projects, the University of Wisconsin set up a separate non-profit facility called the WiCell Research Institute.

This twin-track approach contrasts with the situation in Britain, where both private

and government-funded research fall under a single regulatory framework, with no restrictions limiting researchers to certain cell lines. "The UK has, I believe, sensible regulations," says Anne McLaren of the Wellcome/CRC Institute of Cancer and Developmental Biology in Cambridge, who is a member of the Human Fertilisation and Embryology Authority, the body responsible for granting research licenses.

If ES cells do show clinical promise, and US researchers start falling behind because of restrictions on their access to the leading cell lines, Bush will come under pressure from researchers and patients' groups to amend the regulations. "As soon as we have some evidence of the potential of ES cells, the whole issue can be revisited," says Ron McKay, who works on stem-cell therapies for Parkinson's disease at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland.

In that regard, researchers are encouraged by recent polls showing that US public opinion is strongly in favour of ES-cell research. Many attribute this to media coverage of the subject — most of which has stressed the potential benefits of the research. "I think it is great that the debate brought ES-cell research to public attention," says Peter Mombaerts, who works on mouse ES cells at Rockefeller University in New York.

But any future extensions to the list of cell lines available to federally funded US researchers would require Bush to break a pledge made to his supporters on the religious right of the Republican Party — he has said that he will not change the terms of the announcement he made on 9 August. "If the line is to be moved it will be moved by another president," says C. Ben Mitchell, senior fellow of the Center for Bioethics and Human Dignity in Bannockburn, Illinois, which opposes human ES-cell research.

So, the stage may be set for further debate and lobbying. In the long run, say stem-cell researchers, ensuring the availability of US federal funds for cutting-edge ES-cell research is about more than promoting the competitiveness of US biomedicine. Given the huge investment that may be necessary to bring ES-cell therapies to the clinic, the involvement of the NIH, with its multibillion-dollar budget, may be crucial. "It is absolutely important that NIH researchers work on any cell lines that are developed," concludes Juttner of BresaGen. ■

Laura Bonetta is a writer in Washington.



Dark age: Jim Clark warns that US biomedicine will fall behind.

As time goes on, there will be more and more frustration among American scientists



Bated breath: a Parkinson's disease patient listens to Bush's stem-cell announcement.

Need for public debate about fertility treatments

Manipulations of the mitochondrial germ line must be openly debated and followed up.

Sir—Barritt *et al.*¹ have reported the birth of several 'transmitochondrial babies' following ooplasmic transplantation at *in-vitro* fertilization (IVF) centres around the world. In this technique, about 10% of the ooplasm is transferred from a healthy donor egg into a compromised patient's egg². Because normal human oocytes contain between 200,000 and 5 million mitochondria, the transferred ooplasm probably contains about 20,000–500,000 mitochondria and mitochondrial (mt) DNAs, together with other cytoplasmic components.

Ooplasmic transplantation involves the mixing of mtDNAs in the cytoplasmic germ line. Because this may produce durable and unknown genetic effects that transcend single generations, this new technology must be scrutinized both publicly and scientifically. The social, medical and legal questions about the birth of children who bear the DNA of three parents — mtDNA from the biological mother and the ooplasm donor, and nuclear DNA from both biological parents — cannot be fully anticipated without public discussion and timely scientific publication of results. Cohen and his colleagues are to be commended for their open reporting of rare cases of aneuploidy and other complications associated with ooplasmic transplantation³.

We agree with Parens and Juengst⁴ that increased public supervision is needed. Along these lines, the US Food and Drug Administration (FDA) issued a letter on 6 July 2001 (www.fda.gov/cber/ltr/cytotrans070601.htm) to investigators using ooplasmic transplantation and other forms of gene transfer in human gametes. Investigators using these technologies are now required to submit an 'investigational new drug' (IND) application to the FDA, outlining the research methods and objectives of the clinical study.

We wish to make two additional recommendations. First, we suggest that the National Institutes of Health (NIH) expands its Recombinant DNA Advisory Committee, or creates a new regulatory body, to encompass all ooplasm and nuclear-transfer technologies, to allow informed regulation and public reporting. Until a more solid public and scientific database is established, the number of licences issued should be limited to a handful of highly qualified, excellent IVF facilities. Each of these should be required to follow certain protocol-driven data collection and publication procedures.

Data collection would include haplotype analysis of donor and recipient mtDNAs, screening for the most common pathogenic mtDNA mutations, documentation of gestational complications and maternal–fetal immunological and metabolic parameters, and careful neurodevelopmental testing of the infants and toddlers. The NIH regulatory body should include: people with expertise in a wide range of IVF technologies, including ooplasmic transplantation; specialists in medical ethics, nursing, law, mitochondrial cell biology, mitochondrial medicine and genetics; and informed members of the general public.

Our second recommendation is that the US Congress set aside a budget for a request for applications, issued by the NIH, to solicit proposals to study animal models of the ooplasmic transfer technology. Although significant biological differences make direct comparison of the details of reproductive cell biology between species difficult, these

studies would provide the basic data needed to understand the pros and cons of ooplasmic transfer technology more fully⁵. It is only with carefully regulated progress, timely and systematic publication, public supervision and public debate that we can maintain public trust in science, while moving forward in the genomic and post-genomic ages of medicine that lie ahead.

Robert K. Naviaux*, Keshav K. Singh†

*The Mitochondrial and Metabolic Disease Center, University of California, San Diego, 214 Dickinson Street, CTF, C-103, San Diego, California 92103-8467, USA

†Johns Hopkins Oncology Center, Bunting-Blaustein Cancer Research Building, 1650 Orleans Street, Room 1-143, Baltimore, Maryland 21231-1000, USA

1. Barritt, J. A., Brenner, C. A., Malter, H. E. & Cohen, J. *J. Hum. Reprod.* **16**, 513 (2001).
2. Cohen, J., Scott, R., Schimmel, T., Levron, J. & Willadsen, S. *Lancet* **350**, 186 (1997).
3. Barritt, J. A., Brenner, C. A., Willadsen, S. & Cohen, J. *J. Hum. Reprod.* **15**, 207 (2000).
4. Parens, E. & Juengst, E. *Science* **292**, 397 (2001).
5. Thorburn, D. R., Dahl, H.-H. M. & Singh, K. K. *Mitochondrion* **1**, 123 (2001).

Risk that websites could break code of anonymity

Sir—Anonymous peer review is the fundamental process by which the scientific community evaluates grant proposals. One way in which the Internet has changed the review process is that granting agencies now allow, and some encourage, the construction of websites by authors to supplement proposals. Although these sites provide useful methodological and other information to reviewers, they threaten reviewer anonymity.

Server software can log the unique IP address of each computer that visits a particular website. These IP addresses can identify the institution and, in some cases, the name of the visitor. This is especially problematic when reviewers are the only visitors to a site created specifically for the proposal and hosted by the author, who has disclosed the URL only in the proposal. Reviewers who are aware of the risk of exposing their identity in this way may avoid websites associated with proposals, effectively rendering their review incomplete.

We contacted several funding agencies (NSF, USDA, NERC and the Wellcome Trust), and none has a policy regarding the use of websites in conjunction with grant

proposals, nor do they have mechanisms to protect the anonymity of reviewers visiting proposal websites.

In the short term, we suggest that reviewers visit such sites using a dial-up Internet service provider (for example, America Online and NetZero) that will not provide any information about the reviewer's identity. As a long-term solution, granting agencies should consider hosting such websites rather than allowing submitting authors to host them at their home institution.

Whereas this Correspondence is primarily concerned with the grant review process, the same problem applies to articles submitted to journals for publication, if supplementary material essential to evaluate the article is supplied for peer-reviewers at an author's own website.

David L. Reed, Brett Moyer, Dale Clayton

Department of Biology, University of Utah, 257 South 1400 East, Room 201, Salt Lake City, Utah 84112, USA

Nature has always required supplementary information to be submitted to our editorial offices for peer-review and editorial evaluation. We do not allow such material to be hosted on an author's own site. See guide to authors at <http://www.nature.com/nature/submit/gta/index.html#4.10> — Editor, Nature

IVF and the history of stem cells

Embryo stem cells are poised to fulfil their considerable historical potential.

R. G. Edwards

On the verge of clinical application, stem cells offer a startlingly fundamental approach to alleviating severe incurable human maladies. Fondly believed to be a recent development, they have in fact been part and parcel of human *in-vitro* fertilization (IVF) from as long ago as 1962.

Stem cells arose from attempts made in those early days to obtain cell outgrowths — from ova, cleaving embryos or blastocyst inner cell mass. Grafts of fetal and adult tissues into deficient recipients, including humans, had been used for many years, but nothing about embryo stem cells was known at that time.

Origins

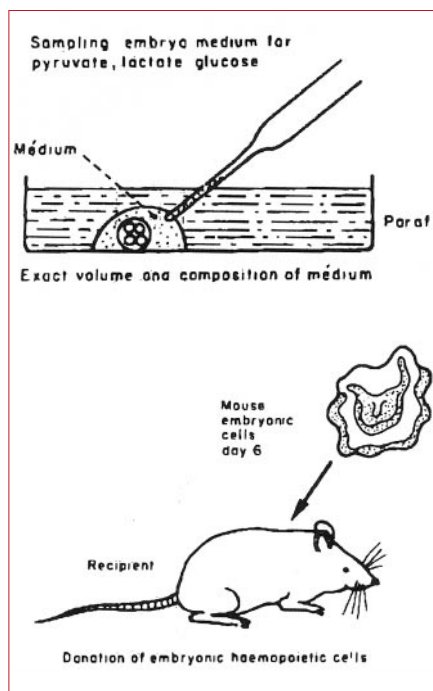
The stem-cell story began when, aware of early work by Maximow¹, we decided to study differentiation *in vitro* of cells disaggregated from mammalian embryos at the stage of development before their implantation into the uterus².

The properties of these pre-implantation blastocyst embryo stem cells were astonishing (see Box 1). All the cell lines possessed large nuclei and distinct nucleoli, and retained typical morphological, karyotypic and enzymatic characteristics throughout successive generations. Long-lived and stable, they differed radically from the typical somatic cell lines with their limited lifespans³.

Stimulated by these findings, I encouraged my Cambridge graduate students to consider working with stem cells for their doctoral research. Richard Gardner tested the insertion of these cells into the mouse blastocoel cavity to discover if they would join with the inner cell mass to co-colonize organs in the resulting chimaeras⁴. Gardner made this idea work, producing the world's first mouse chimaera, its typical coat-colour pattern becoming the instantly recognized symbol for transgenic mice.

For me, human IVF was moving fast and I had to place the stem-cell projects on the back burner. The concept of embryo stem cells transferred to the Department of Genetics at Cambridge University, where much elegant and well-known work was done by many scientists (see ref. 5, for example). 'ES' cells became a familiar term, reflecting their immense potential in colonizing chimaeras.

As the work continued and IVF developed, we grew the first human blastocysts *in vitro* at 5 days after insemination, thus overcoming the previous near-impossibility of



Early days: how mouse embryos could be used to repopulate the haematopoietic system, drawn in 1986 by Peter Hollands.

obtaining human blastocysts⁶ (see Box 2).

The distinct inner cell masses and — even more encouragingly — the large embryonic

discs in blastocysts aged 9.5 days shouted news of embryo stem cells. Some years later, the birth of children conceived *in vitro* confirmed that human blastocysts *in vitro* are capable of normal growth, offering a reliable source of inner cell mass.

Ideas for using human stem cells therapeutically re-emerged, but of course animal studies were an essential first step. Another of my doctoral students, Peter Hollands, investigated whether mouse embryo stem cells could colonize and repair damaged tissues in adult recipients, using the restoration of bone-marrow function in X-irradiated or severely anaemic mouse recipients⁷.

The idea of this model was to avoid the graft-versus-host responses that plagued many other approaches to therapies, and make available 'space' in depleted bone marrow for colonizing stem cells. Hollands achieved his target, as disaggregates of numerous mouse blastocysts or of 7-day-old fetuses injected into recipient tail veins conferred a full lifespan; controls died in a few days. Chromosome markers revealed grafted mouse or rat embryo stem cells successfully colonizing bone marrow after a few days, and biochemical markers identified donor red cells and lymphocytes in recipients. Hollands did not observe graft-versus-host responses, but noted suggestions of natural repair.

More astonishing still, I scored Hollands'

Box 1 First experiments

In the 1960s, we found that cell lines from individual cells of rabbit cleavage-stage embryos grew only sporadically in various media, even when we included feeder cells. Blastocyst cells differed, and provided the basis of most subsequent work to the present day. Cultures of intact rabbit blastocysts aged 5.5 days, sometimes in media reinforced with 'mesodermalizing inductors' and occasionally with feeder layers, yielded cell clumps, sheets and single-cell outgrowths. Inner cells migrated over a forming pavement of trophectoderm to form various structures and individual cells *in vitro*. Outgrowths contained blood islands, muscle, connective tissue, neurons and macrophages among other undefined tissues, revealing the breadth of tissues easily available from inner cell mass². Among them, haematopoietic precursors were most expected, as blood islands form *in vivo* a few days post-blastocyst, as in humans.

In a different approach, we dissected whole rabbit inner cell masses from 6-day-old blastocysts and placed them on collagen-coated coverslips in droplets of medium, their inner surfaces facing collagen, to which they attached after 10–15 hours². Multicellular tissue sheets formed, fibroblastic and epithelioid migratory monolayers covered the collagen surface, and disaggregated embryonic discs of 6.5-day-old blastocysts, completely coated collagen-coated coverslips in a day or two. Some colonies later attached directly to glass, enabling cell lines passaged by trypsinization to be established. We could derive 20 subcultures from a single disc, many flourishing *in vitro* as distinct cell lines displaying immense properties of long-term survival ('immortal', in today's terminology). Either approach, using whole embryos or cell lines, promised remarkable therapies.

Two epithelial-like cell lines passaged from 6 weeks after explantation included one with high alkaline phosphatase, which grew vigorously in numerous passages over 9.5 months and remained largely diploid. A fibroblast strain rich in arginase grew continuously for 11 months through 40 subcultures (200 generations), again remaining largely diploid. When this line was accidentally lost, we could restore this activity from cryopreserved cells².

coded microscope slides to reveal donor mouse and rat stem cells migrating through mouse liver to bone marrow, apparently along pre-established fetal pathways in recipients⁸. We assumed that these cells were migrating haematopoietic cells, but they might have been donor hepatic cells. Either way, the result was incredible, now confirmed by Catherine Verfaillie (University of Minnesota Cancer Center; personal communication, 2001), who identified the route as through lung and liver to bone marrow in recipients. Hollands⁸ could also weakly colonize non-irradiated recipients by stem cells that were not rejected, in contrast to skin grafts of the same genotype. *Nature* was unimpressed, rejecting our paper — which certainly had flaws, but not enough of them to discard a stunning new idea.

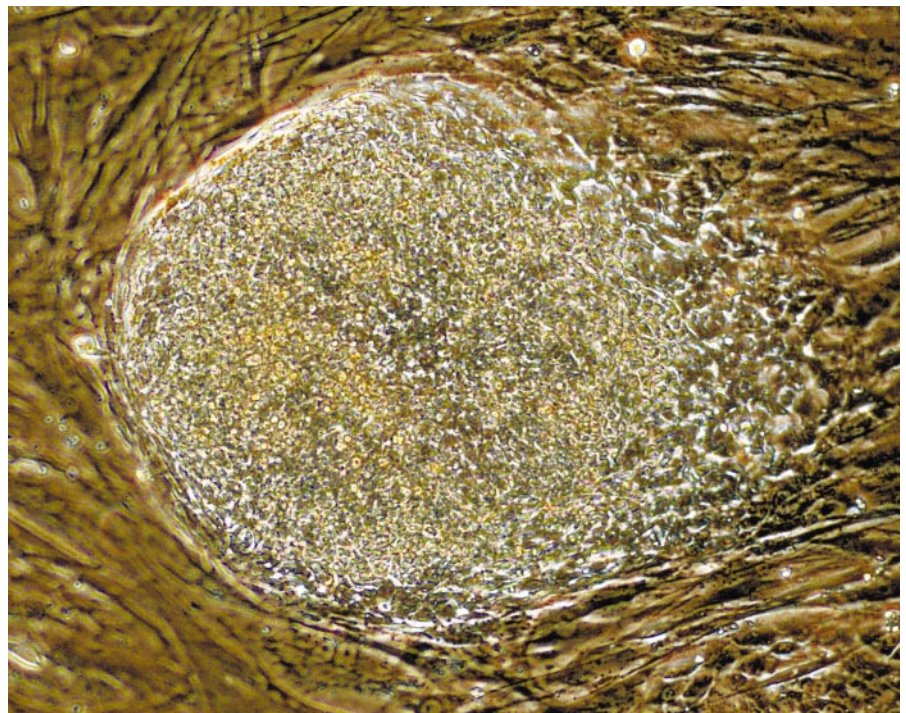
Exciting prospects

What an exciting prospect of embryo stem cells had emerged! The cells were long-lived ('immortal'); stable enzymatically, karyologically and morphologically; capable of very wide differentiation patterns; able as single cells to colonize vast areas of chimerae; capable of repairing tissues in sick recipients; of following pre-existing fetal pathways of tissue colonization; persisting *in vivo* throughout the lifespan; species non-specific; and partially free of host-versus-graft or graft-versus-host reactions. Haematopoietic stem cells found their own way to target tissues.

In successive articles and meetings, I presented concepts on colonizing sick human organs, or providing new stem cells after the originals had been used^{9,10}. Just as I had experienced with IVF and pre-implantation genetic diagnosis, many people dismissed the concept as far-fetched. Yet today I delight in seeing other investigators displaying the same euphoria, rediscovering the simplicity of using stem cells and their immense benefits.

Very rare human blastocysts became available in my laboratory at Bourn Hall in the early 1980s, enabling us to try to produce human stem cells. Whole embryos cultured *in vitro* produced outgrowths that lived for several days and secreted human chorionic gonadotrophin, but they were not suitable for stem-cell outgrowths. *Science* published our initial paper¹¹, but reluctantly and with cutting remarks about underlying ethical objections to this use of human embryos. When they were finally produced, human stem cells would require isolated inner cell masses for cultivation on collagen, as in our earlier work on rabbits².

Unfortunately, my work on stem cells ended with an ethical decision at Bourn Hall, which was taken with my support, to cryopreserve all excess embryos for their parents. From then on, the only source of blastocysts would be those that had been cryopreserved



Exciting developments: a human stem-cell colony growing on mouse embryonic fibroblasts.

but were no longer wanted by parents, which lay several years away.

Legalities and ethics

Rightly, ethical concern has been paramount in regulating research on assisted human reproduction and on human embryos. Many of the embryos we had cryopreserved, imagining them as a reserve for future research with their parents' permission, were destroyed for legal reasons in the United Kingdom in the 1990s. Contact with their parents had been lost, so there was no other option.

It is ironic that in the United States today, President George W. Bush and the National Institutes of Health object ethically to using human embryos to make stem cells, yet accept 64 cell lines prepared by foreign investigators or by private US clinics (see page 345)!

Embryos are not chattels, and doubts will always exist about parents' rights to donate them for research. Buying or using stem cells from someone else to keep one's own conscience clear is just as unethical as the original decision to make them, or more so. It is as deceitful as a motion raised (but not formally proposed) for the European Parliament to ban EU funds from clinics practising research on human embryos, while the would-be proposers simultaneously enjoy the fruits of such research for their citizens.

Nor can ethical committees, the UK's Human Fertilisation and Embryology Authority (HFEA) or governments give consent, for the same reason. They can waive prosecution for individual scientists doing research who, driven by the clinical impera-

tive, decide personally to proceed. This is difficult ethical territory, with its overtones of complicity, yet this is how almost every recent advance in reproductive technology began, and it is what is happening now worldwide with human stem cells.

Some investigators are now patenting their findings. Surely, patents should be withheld in such ethical situations, or at least in those cases where the first ideas came from freely available, published, basic animal work¹² (which is often unquoted in patent applications).

The current ethical furore cannot hide the fact that stem cells receive top-level encouragement, which is not offered for IVF. Even human cloning, the worst concept imaginable to many people, has been encouraged by the UK government and the Royal Society, whose motivation was to allow research to eliminate recipient graft rejection of stem cells. Predictably, the decisions they made in this area were taken as encouragement by those wishing to make human clones.

As things stand, three-quarters of human embryos (*in vivo* and *in vitro*) are unable to implant, dying before or very soon after this event. Cloning could make this situation worse. How would stem cells perform under such circumstances? Cloning may not even be necessary in order to avoid rejection, as embryo stem cells weakly avert rejection⁹, and mild immunosuppressives confer some protection¹⁰.

Yet some commentators imply that today's higher ethical standards would have precluded human IVF. This belief lies far from reality. IVF is perhaps the most

regulated clinical speciality worldwide, as governments everywhere dictate what can be done in this field.

No US president shaded his beliefs, and no UK prime minister introduced the most controversial ethical constraint of all, to make embryo stem cells. Stem-cell ethics are minor when compared with the issues raised by gamete donation, surrogate pregnancy, embryo cryopreservation, research on early human embryos, pre-implantation genetic diagnosis and adult cloning.

Potential of ES cells

Fortunately, such troubling ethical problems do not deter investigators from clarifying the immense potential of animal and human stem cells (see ref. 13 for a recent summary).

The widening scope of animal studies points the way to the use of human embryo and tissue stem cells. Those prepared from embryonic tissues, primordial germ cells, early and late fetuses, and cord blood are multipotential, exhibiting remarkable feats of rapid colonization and differentiation of haematopoietic, neural, skeletal, muscle and other tissues. Reported recoveries from serious injury are remarkable, promising immense advantages from the 64 human stem-cell lines identified by President Bush.

There are four key publications on human stem cells that demonstrate the potential of this field. The first of these was the initial report of human stem cells¹⁴, which were isolated from blastocysts and grown with mouse feeder fibroblasts to produce gut epithelium, cartilage, bone, muscle, ganglion and other cell types after 4–5 months, using methods virtually identical to

Surely, patents should be withheld where the first ideas came from published, freely available work.

ours in rabbits 30 years earlier². The human cells were characterized by distinct nuclei, high telomerase levels, and surface markers typical of undifferentiated ES cells.

In another study, human primordial germ cells cultured *in vitro* retained their karyotype while producing large, ES-like cell colonies capable of repeated passages¹⁵. In the third, markers ensured that the correct lineages had been established, the results indicating that specific media hold promise for directed differentiation of cells¹⁶.

In the fourth study, eight growth factors were tested on a human ES cell line; the cells expressed receptors for each growth factor. The three germ layers, as well as eleven tissues, were identified by molecular markers, and specific gene expression in germ layers and tissues was identified by polymerase chain reaction¹⁷.

These studies show that the time is rapidly approaching when close control over directed differentiation of stem cells into each germ layer, or into specific tissues, will become possible.

All this knowledge, and more, predicts an astonishing future. It was clearly no coinci-

dence that, in the early days, so many tissues differentiated in our outgrowths of rabbit inner cell mass or that mouse stem cells migrated through liver to bone marrow. Local guidance cues and established pathways apparently enhance an enormous expansion of stem cells at sites of injury and repair, with risks of tumours and graft rejection being low. Patients may soon be helped by the use of small numbers of specified stem cells, and vats containing millions may become available for large-scale grafting or to make products, just as is now the case with recombinant gonadotrophins.

Next steps

The fundamental questions remaining to be answered include how and where these complex cells originate, and how they distribute so widely in tissues, even in cord blood. Are they close to primordial germ cells, having switched to a repair function during early embryogenesis rather than undergoing meiosis in specialized gonads? Or do they specialize independently during early differentiation, as a sort of diffuse repair system typical of each tissue?

Such questions need answers, and early human differentiation must be better understood¹⁸, to improve endogenous forms of tissue repair. ■

R.G. Edwards is Editor of Reproductive BioMedicine Online, Duck End Farm, Dry Drayton, Cambridge CB3 8DB, UK, and winner of the 2001 Lasker Award for clinical research (see Nature 413, 242; 2001).

1. Maximow, A. *Contrib. Embryol. Carnegie Inst. Washington* **16**, 47–113 (1925).
2. Cole, R., Edwards, R. G. & Paul, J. *Dev. Biol.* **13**, 285–407 (1967).
3. Hayflick, L. & Moorhead, P. S. *Exp. Cell Res.* **25**, 581 (1961).
4. Gardner, R. L. *Nature* **220**, 596–597 (1968).
5. Evans, M. J. and Kaufman, M. H. *Nature* **292**, 154–156 (1981).
6. Edwards, R. G. *Conception in the Human Female* (Academic, London, 1980).
7. Hollands, P. *Development* **99**, 69–76 (1987).
8. Hollands, P. Br. J. *Haematol.* **69**, 437–440 (1988).
9. Edwards, R. G. (ed.) *Fetal Tissue Transplants in Medicine* (Cambridge Univ. Press, Cambridge, 1992).
10. Edwards, R. G. in *Human Contraception in Vitro* (eds Edwards, R. G. & Purdy, J. M.) 371–388 (1982).
11. Fishel, S., Edwards, R. G. & Evans, C. J. *Science* **223**, 816–818 (1984).
12. Biggers, J. *Reprod. BioMed. Online* **1**, 74–76 (2000).
13. The Royal Society. *Stem Cell Research and Therapeutic Cloning: An Update* (Royal Society, London, 2000).
14. Thomson, J. A. et al. *Science* **282**, 1145–1147 (1998).
15. Shambloot, M. J. et al. *Proc. Natl Acad. Sci. USA* **95**, 13, 726–731 (1998).
16. Pittenger, M. F. et al. *Science* **284**, 143–147 (1999).
17. Schuldiner, M. et al. *Proc. Natl Acad. Sci. USA* **97**, 11, 307–312 (2001).
18. Edwards, R. G. *Reprod. BioMed. Online* **3**, Webpaper 2001/156 (2001).
19. Edwards, R. G. *Lancet* **2**, 926–929 (1965).
20. Edwards, R. G., Bavister, B. D. & Steptoe, P. C. *Nature* **221**, 632–635 (1969).
21. Steptoe, P. C., Edwards, R. G. & Purdy, J. M. *Nature* **229**, 132–133 (1971).
22. Edwards, R. G. & Surani, M. A. H. *Uppsala J. Med. Sci. Suppl.* **22**, 39–50 (1978).
23. Edwards, R. G., Steptoe, P. C. & Purdy, J. M. *Brit. J. Obstet. Gynecol.* **87**, 737–756 (1980).
24. Steptoe, P. C., Edwards, R. G. & Purdy, J. M. *Brit. J. Obstet. Gynecol.* **87**, 757–768 (1980).

Box 2 Human IVF and stem cells

Embryo stem cells developed from the beginnings of IVF research, when oocytes were matured and fertilized *in vitro*^{19,20}. Moving to clinical work involved aspirating fully mature human oocytes from patients' ovaries after mild gonadotrophin stimulation²¹. These oocytes could be fertilized *in vitro*, and grew through cleavage stages to blastocysts after 5 days in culture, and then to 9 days²² with a large embryonic disc. Embryo transfers to the uteri of infertile patients were initially unsuccessful, but improved luteal phase support or the use of natural cycles led to pregnancies and full-term babies^{23,24}. Human blastocysts cultured *in vitro* clearly had the same developmental possibilities as those growing *in vivo*, enabling work to begin on human stem cells for therapeutic purposes.

Chronology

- 1962 Systematic collection of mature oocytes from patients' ovaries
- 1962 Rabbit embryo stem cells grown and differentiated *in vitro* from inner cell mass
- 1963 Rabbit stem-cell lines established
- 1965 Human oocytes matured *in vitro* to metaphase 2 and first polar body
- 1967 Cells injected into mouse blastocysts produce mosaic chimaeras
- 1968 Rabbit blastocysts sexed successfully
- 1969 Fertilization *in vitro* of human oocytes matured *in vitro*
- 1970 Mature oocytes aspirated from human follicles after gentle priming of patients with gonadotrophins
- 1971 Human fertilization and embryo growth *in vitro* to expanding blastocysts
- 1972 Human embryo transfers begin
- 1977 Human embryos grown to day 9, with huge embryonic discs trumpeting news of stem cells
- 1978 Birth of first IVF baby
- 1984 First attempt at producing human stem cells

Does ivory burn?

One man's battle to protect Africa's elephants from extinction.

Wildlife Wars: My Fight to Save Africa's Natural Treasures

by Richard Leakey & Virginia Morell

St Martin's Press: 2001. 352 pp. \$25.95

Colin Tudge

Richard Leakey is, at the very least, an extraordinary person. Born in 1944 to the great East African palaeoanthropologists Louis and Mary Leakey, he shunned his parents' trade and left school at 16 to start a safari business. But in 1963 he learned to fly ("to cure my fear of flying"), and on his first solo outing he saw deposits around Tanzania's Lake Natron that reminded him of Olduvai, and decided that he, too, wanted to dig. Thus, though without university education (which continues to bother him), he came to dominate the search for fossil hominids in East Africa, not least as director of the Kenya National Museum in Nairobi.

Along the way he was diagnosed with terminal kidney disease, to be rescued in 1979 by a transplant from his brother Philip. He also chaired the board of the East African Wildlife Society and so was introduced to the plight of elephants. And with characteristic shoot-from-the hip opportunism and courage he engineered a press conference on ivory poaching at the Nairobi museum (at a meeting ostensibly about birds). So, "out of the blue", in April 1989 President Daniel Arap Moi appointed him to be director of Kenya's Wildlife Department — later renamed the Kenya Wildlife Service, or KWS. Thus begins his present tale: a unique, frank and terrifying insider's view of conservation, of a developing country and, indeed, of humanity in the raw.

For, as emerges from incident after bloody incident, "Kenya's politics are rough". Kenya has no minerals to speak of, or oil. Tourism is the greatest earner, and at the heart of it is wildlife and in particular the elephants. In crude economic terms the elephant is to Kenya what diamonds were to South Africa and oil should have been to Nigeria. But when Leakey took over at the Wildlife Department, corruption ran from the local shopkeepers to the heights of government and the elephant was heading pell-mell for extinction. Kenya had 85,000 elephants in 1979; in 1989, just 22,000.

Trade in ivory was legal. But the poaching was much bigger, aided in large part by the Wildlife Department itself, with elephants gunned down in full view of tourist lodges. The honest souls were on their beam-ends. Leakey's first inspection of his men revealed downcast rows of wardens with no boots and First World War rifles that jammed after the

first bullet, who set off in broken-down landrovers to take on poacher-guerillas armed with modern assault rifles of the kind that, in modern Africa, are all too easy to come by. He put his own name on cheques to re-equip his men while raising money from abroad, and answered the poachers' open mockery in kind. When they left him a note in a tree, "We will kill you, Leakey", he wrote on the back: "Perhaps. But my men will kill you first." This was street-fighting, but appropriate.

The legal trade in ivory was supposed to raise revenue for conservation. But although Leakey acknowledges that wildlife must help to pay for itself he also believes that whereas tourism is fine and necessary, trophies are not. He is a pragmatist first: "The more I found out about the ivory trade, the less sure I became that there was a way to manage it legally", and, "As long as ivory was considered a valuable commodity, there would be a black market for it".

But there's a strong moral thread, too. Leakey was introduced personally to elephants by Joyce Poole, the kind of modern biologist who, like Jane Goodall and George Adamson, forms close relationships with wild animals that for other people would be lethal. To Leakey's horror, not to say stark terror, she drove them to the middle of an elephant family and proceeded to commentate on their feelings and to predict how each would behave. The point is not simply that elephants are 'intelligent', in a maze-solving way. It's their awareness that startles: their sensitivity and family feeling. To poach them looks very like murder.

So then, famously, in 1989 Leakey held up the legal sale of 2,000 tusks on a bureaucratic

technicality and, although they were worth at least US\$3 million, resolved to burn them. Meave, his wife, was supportive as ever, and so, to his eternal credit, was Moi. But both asked the obvious question: does ivory burn? Meave and Leakey tried a bit on the fire, and it didn't. But a friend of a friend who did special effects in Hollywood said the tusks would burn if coated with flammable glue and fired with truly nebuchadnezzar intensity.

Thus, on 18 July 1989, Moi made a short but statesmanlike speech and applied the flame to the 20-foot tower of incendiary tusks on bales of straw that had been soused in petrol and kerosene under pressure. Moi still feared a fiasco — that the tusks would stay intact. Leakey feared that the entire assembly, TV crews and all, would go up in smoke. In the end, the blaze was fierce but safe, the tusks were reduced to pale grey ash, and the event was witnessed by 850 million people worldwide.

Leakey then strove for a worldwide ban on the ivory trade, and more rough dealings followed with the southern Africans, who favoured controlled trading. At the 1989 meeting of the Convention on International Trade in Endangered Species (CITES) in Kyoto, the "beefy" Rowan Martin of Zimbabwe National Parks called Leakey a "bunny hugger". Leakey says he replied with "what I hope was withering sarcasm" — so much so that he embarrassed his own side. His conservationist opponents, he says, "were not people I would have invited home". But he won the day. Elephants were given CITES Appendix I status in October 1989, effectively banning trade. And although the battle goes on, the price of ivory dropped from \$100 per pound to less than \$5.



ALL IMAGES BY DAVID NEWTON

Life then went from rough to rougher. No one knows whether the plane crash in June 1992 was sabotage, but Leakey lost both his legs (managing to insult Moi, who had hurried to his bedside, by telling him he didn't need his prayers). He was forced to resign from KWS in March 1994, started his own political party in opposition to Moi and became a member of parliament in 1997 (and campaigned for the disabled). But in 1998 he was reinstated — by Moi — to direct KWS afresh. Conciliation, it seems, is a vital part of African politics. Now he has left KWS a second time to become permanent secretary to the president and head of Kenya's Public Service.

All in all, this is a rattling good yarn. But it is much more than that. At stake is the concept of nationhood, a bureaucratic artifice superimposed on tribes that are ancient, evolved and real. The battle of tribal and political leaders echoes that of Europe's medieval barons, albeit fought out under modern political floodlights. Greater than that is the perennial struggle of humans to form a coherent society, and although most are honest and socially minded, the influence of those bent only on self-interest is inevitably disproportionate. In the background is humanity's attempt to find a *modus vivendi* with a fellow creature that needs to be managed, but is sensitive beyond present measure and must not be 'controlled' as if it were some pest.

How will Leakey himself be remembered? He says he did not expect to survive beyond 40 but as he now seems indestructible, it is clearly too early to say. Colonial and post-colonial Africa has been shaped by larger-than-life personalities, some mercifully great, such as Nelson Mandela and Desmond Tutu, but also by too many chancers, such as Cecil Rhodes. Leakey has his detractors, but if we acknowledge that all humans are flawed, then he surely belongs among the greats. Kenya's elephants would sign up to that. ■

Colin Tudge is a writer based in Oxfordshire, UK.
His latest book is *In Mendel's Footnotes: An*

Introduction to the Science and Technologies of Genes and Genetics From the 19th Century to the 22nd (Jonathan Cape).

Wormholes through physics

Nine Crazy Ideas in Science: A Few Might Even Be True

by Robert Ehrlich

Princeton University Press: 2001. 254 pp. \$24.95, £16.50

Time Travel in Einstein's Universe: The Physical Possibilities of Travel Through Time

by J. Richard Gott

Houghton Mifflin: 2001. 291 pp. \$25, £18.99
Paul Davies

Niels Bohr once remarked, when confronted with an unorthodox theory, that the problem was not whether it was crazy, but whether it was crazy *enough* to be true. Modern science, especially physics, is replete with outlandish ideas that defy common sense and intuition. It is almost impossible for the non-scientist to discriminate between the legitimately weird and the outright crackpot. If black holes, antimatter and virtual quantum particles are taken seriously, why is faster-than-light travel or extra-sensory perception pooh-poohed by scientists?

Physicist Robert Ehrlich has assembled a fascinating collection of apparently crazy ideas, and subjected them to careful analysis, assigning each a 'cuckoo' index of implausibility. Some of them, like the claim that small doses of nuclear radiation might actually be beneficial, he finds fairly convincing. To others he gives a decisive thumbs down, such as the much-publicized theory that AIDS isn't caused by HIV, or that having more guns in circulation leads to less crime. Ehrlich points out how statistics can be misleadingly presented in these cases, and how the distinction between effects that are causally related and those that are merely correlated often gets blurred.

Readers will make their own judgements about the various claims. Personally, I would not give the existence of tachyons — hypothetical particles that exceed the speed of light — quite the zero cuckoos that Ehrlich assigns. I do, however, share his enthusiasm for the offbeat theory of Thomas Gold that reserves of oil, coal and gas are not all produced by biological processes (the conventional story) but can be generated by primordial methane percolating up from deep in the Earth's crust and subjected to the activities of subterranean microbes. Gold has received a lot of flak for flying in the face of geological orthodoxy, and although I feel that the case he makes is far from decisive, I believe he

deserves far more credit for his ideas. In particular, his early championing of the theory that there exists a deep, hot biosphere turned out to be startlingly accurate, and is now generally accepted.

One crazy idea to which Ehrlich assigns medium plausibility is time travel. It is a recurring theme of science fiction, but can it really be done? Travel into the future is unproblematic, and known to be possible; it is a standard prediction of Einstein's theory of relativity and has been confirmed experimentally many times. But going back in time is another matter entirely. If that were possible, then all sorts of causal paradoxes loom, such as the famous example of the time traveller deliberately killing his grandfather as a baby, thus negating his own existence.

In spite of its bizarre overtones, the possibility of visiting the past has attracted the attention of many serious researchers in recent years. Most of the work relates to the proposal that a wormhole in space could be used as a time machine. A wormhole is like a black hole, but with an exit as well as an entrance. If one existed, it could provide a short-cut between distant points in space. Whereas falling into a black hole would be a one-way journey to nowhere, entering a wormhole might enable you to come out moments later in another part of the Universe. The theory of relativity predicts that a wormhole can be adapted so that an astronaut traversing one would come out *before* he went in. Wormholes remain highly speculative, but no knock-down argument yet exists that proves them to be physically impossible.

In *Time Travel in Einstein's Universe*, Richard Gott has come up with an alternative method of visiting the past, using hypothetical entities called cosmic strings. These would be exceedingly narrow tubes of trapped energy left over from the Big Bang, containing enough mass to seriously warp space and time. Gott envisages a pair of parallel cosmic strings flying past each other at high speed, and argues that an astronaut following a path looping around the strings could be transported back in time.

Gott's proposal is hardly practicable, being even less plausible than engineering a wormhole. Still, the merit in studying time travel lies not so much with the practicalities, but more with the way in which the topic illuminates the deep foundations of physics. Obviously, science must give a rational and self-consistent account of reality, so genuine paradox is not permitted.

Gott tries heroically to make the subject accessible to the non-specialist reader, although some of the discussion gets pretty technical, and occasionally mysterious. Perhaps his most striking claim is that the entire Universe could be a time machine that reaches back to the Big Bang and engineers its own creation. Some readers may well regard this



idea as a *reductio ad absurdum* of the whole subject, but for those who enjoy a speculative romp across some of the most fascinating topics in modern physics, Gott's book will prove a delightful challenge. ■

Paul Davies is visiting professor of physics at Imperial College, London SW7 2AZ, UK, and author of *How to Build a Time Machine* (Penguin), to be published later this year.

Cut-and-paste knowledge

Encyclopaedic Visions: Scientific Dictionaries and Enlightenment Culture

by Richard Yeo

Cambridge University Press: 2001. 358 pp. £40, \$59.95

W. F. Bynum

One of my son's babysitters was an elderly woman who also looked after children of other Cambridge research students. When she closed her cottage to move in with her own son, she rewarded all her bookish friends with a remembrance: each received a single volume from her set of the *Everyman Encyclopaedia*. I never knew if the fact that I got the first volume was a particular sign of affection, or something rather more random.

Odd volumes of encyclopaedias may contain a lot of useful information, but they are best in complete sets. 'Twas ever thus, as *Encyclopaedic Visions*, Richard Yeo's sensitive and engaging study of Enlightenment encyclopaedias, makes clear. After all, the cycle embedded in the word 'encyclopaedia' carries with it the implication that all human knowledge is connected in some organic way. Most medieval and early modern works within the genre carried a statement or diagram, or both, attesting to such unity.

In spite of this rich historical tradition, modern encyclopaedias emerged only during the Enlightenment. At first glance, the Enlightenment reference book might be thought to be the *Encyclopédie* of Denis Diderot and Jean D'Alembert. Much has been written about the creation, publishing history and influence of this monumental work, whose aggressively secular nature brought both of its editors (and some of its contributors) into confrontation with the authorities in *ancien régime* France.

As Yeo reminds us, however, the *Encyclopédie* began its life as a planned translation of a work produced single-handedly by an English hack, Ephraim Chambers. So successful was Chambers' *Cyclopaedia*, first published in two volumes in 1728, that it turned its obscure author into a successful man of letters and earned him election to the Fellowship of the Royal Society (the eighteenth-



century Royal Society was happy enough to elicit science popularizers to its midst).

Chambers' *Cyclopaedia* is central to Yeo's analysis, and science was at the heart of Chambers' work, as well as that of his successors and imitators. The objectivity of scientific knowledge gave it a special place in Enlightenment culture, raising it above the divisiveness of politics or religion. Science was thus fundamental to the ethos of self-improvement that featured in the rhetoric with which encyclopaedias were announced, advertised and justified. The scientific and technical contributions of Chambers, John Harris, Abraham Rees and the other encyclopaedists of the period were often highly original and much remarked upon.

But if the encyclopaedias were surrounded by an aura of utility and progress, they were, especially in Britain, products of commercialism and profit. In continental Europe, several learned societies undertook the systematic task of producing versions of the encyclopaedic ideal. Most of these projects either foundered or took decades to be realized. The British seemed to vacillate between pride at their own individual enterprise and initiative and envy of the more systematic way in which learned activity was organized on the Continent. Thus, the first edition of the *Encyclopaedia Britannica* (three volumes, 1768–71) was written single-handedly by the Edinburgh printer William Smellie (by his own admission with a paste-pot and a pair of scissors). Its title page, however, announced that it had been produced by 'A Society of Gentlemen'.

Smellie's mode of composition was hardly surprising: no one, even in the Athens of the North, could be expert in all branches of human knowledge. The key to producing a good encyclopaedia was knowing which authorities to summarize. This raised the tricky issue of copyright, even at a time when the legal protection of intellectual property was weak and pirated editions were regularly produced. The first edition of the *Encyclopaedia Britannica* actually produced a lawsuit, partly because of Smellie's liberal use

of the scissors and paste-pot, but mostly because London publishers and booksellers were nervous that this new encyclopaedia on the block originated in Scotland.

The *Encyclopaedia Britannica* not only survived these early rumblings, it actually thrived, and by the time of its third edition (1788–97) it had swelled to 18 volumes and acquired many of the trappings of modernity. Although still responsive to market forces, this *EB* was the work of identified experts. Although it aimed at comprehensive coverage of all branches of human knowledge, its scientific content was still central, science being recognized as the most active and rapidly changing field of knowledge, and therefore the one that dictated the necessity for frequent updating.

Yeo's monograph is solidly historical, but it reminds us that many of the issues of encyclopaedic form and content grappled with during the Enlightenment are still with us. Should encyclopaedias be principally repositories of authoritative knowledge where the learned can refresh their knowledge, or instruments to educate the novice? If science is about the objective understanding of nature, should it be subject to copyright? These and many other themes that Yeo scrutinizes are still current in our age of electronic publishing and the Human Genome Project. ■

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.

The price of success

Science, Money, and Politics: Political Triumph and Ethical Erosion

by Daniel S. Greenberg

University of Chicago Press: 2001. 528 pp. \$35, £22.50

David Dickson

Two characteristics instantly strike any observer of the scientific enterprise in the United States. The first is its vitality and productivity. The second is a level of financial support that, even in difficult economic times, has remained the envy of virtually all other industrialized nations.

A third characteristic, perhaps less immediately obvious but no less striking, is the longevity of its support structures. In other countries, research councils and funding agencies find their names and responsibilities altered as politicians seek to modify their roles in the face of changed economic circumstances. In contrast, the organizational landscape of US science has remained virtually unchanged for the past half-century.

For anyone interested in understanding

at least the second two of these characteristics, Daniel Greenberg's book, the product of almost four decades of close observation of the Washington science scene by one of its most acute analysts and sharpest critics, is essential reading.

Greenberg, a former news editor of *Science*, for many years chronicled in his fortnightly newsletter, *Science and Government Report*, the complex interaction between the scientific community and the political establishment. No one, therefore, is better placed to document how each has successfully managed to meet the needs of the other since the end of the Second World War.

Greenberg does not deny that this symbiotic relationship must take much of the credit for the current strength of US science and its dominance of the world stage. But he argues that there has been a heavy price to pay in terms of the intellectual and ethical cost of getting there.

For example, one consequence of the successful pursuit of self-interest by the scientific community — the “political triumph” referred to in the book's title — has been an innate conservatism that Greenberg claims has, ironically, led to its increasing estrangement from mainstream politics.

A second consequence, he argues, has been a lowering in the intellectual integrity of the political discourse around science. Greenberg points out, for example, that many still invoke the name of Vannevar Bush to justify the protection of basic research funding from direct political interference. But few of these, he claims, are aware of the extent to which Bush's ideas were modified — and some of them rejected — during the process of setting up the system for funding science after the Second World War with which his name is identified.

At the same time, Greenberg argues, scientists have become increasingly disinclined to take up the moral causes that fired a previous generation of scientific leaders to campaign against issues such as nuclear-weapons testing or environmental degradation. He describes the main preoccupation of today's scientist, outside his or her scientific work, as “grubbing for money”, and laments that “the demobilisation of science from politics and social engagement is a fact of scientific life”.

Greenberg is at his best when tracking in painstaking detail, often using internal documents obtained in the course of preparing material for his newsletter, the way in which dubious lines of argument can take on a virtually unquestioned life of their own if they are found suitable for building a case for greater science funding.

Such, for example, was the case when the National Science Foundation set out in the mid-1980s to argue the case, based on a remarkable lack of hard data, that the United



States faced a damaging shortfall in the production of scientists and engineers. The argument was dropped in the early 1990s when it became clear that this was unlikely to occur.

He is also adept at undermining some of the myths about the degree of influence that scientists have over political affairs outside their direct spheres of activity. Their impotence is reflected, for example, in Greenberg's documentation of the continuous failure of the scientific community — or the “scientific enterprise”, as he calls it — to establish a strong scientific presence in the Department of State. As a result, he writes, the department “has persisted in a benighted indifference to things scientific, sometimes to the astonishment and dismay of scientists who cross its path”.

Many scientists continue to believe that science's generous support from the federal government is based primarily on the innate value of its potential contribution to social well-being. Greenberg's analysis of such events may well cause them to reconsider their view of how decisions about science funding are taken in practice.

There are shortcomings in his analysis that will no doubt be eagerly leapt on by those reluctant to accept the relatively unflattering portrait of US science that Greenberg presents. One is that, for all its institutional conservatism and self-serving politicking, US science has been remarkably productive in the period he describes. As far as science itself is concerned, and whatever Greenberg says about its innate conservatism, it is difficult to see how different strategies could have led to even greater achievements.

Second, there are some significant gaps in the analysis. For example, Greenberg spends considerable time analysing the misguided hubris that accompanied the collapse of political support for the Superconducting Supercollider in the early 1990s — and the political factors that ensured the survival of a

far less scientifically deserving project, the International Space Station.

But he pays scant attention to perhaps the greatest single science project of the past decade, namely the sequencing of the human genome. This receives merely a passing comment, being linked dismissively with various ill-fated experiments in gene therapy as a product of the joint pursuit of “scientific glory and biotech profit”.

Finally, Greenberg's book does little to explain how his main remedy for the ills he describes — that scientists should “come out of the ghetto” and become more directly involved in conventional politics — are likely, on their own, to change things.

None of these shortcomings, however, detracts from the value of this book as a unique and revealing perspective on the way that the science-funding process actually works in Washington. The picture it paints is not a flattering one. But — unlike many of those he writes about — Greenberg is not out to make friends in high places. ■

David Dickson, a former news editor of *Nature*, is at SciDev.Net, 11 Rathbone Place, London W1T 1HR, UK.

An astronomical adventure story

Beyond Pluto: Exploring the Outer Limits of the Solar System

by John Davies

Cambridge University Press: 2001. 244 pp.

£17.95, \$24.95

Joel Wm. Parker

The fact that our Solar System consists of more than just nine planets and an asteroid belt between Mars and Jupiter is still sinking into the public psyche. But among Solar System researchers, the importance of the Kuiper belt beyond Neptune has been known for quite a while. This region contains another ‘asteroid belt’, which is the source of comets and will provide a glimpse into both the chemical and the dynamic infancy of the Solar System.

Considering the recent public debate over the planethood of Pluto, and the on-again, off-again politics of NASA's Pluto–Kuiper Express mission, now is a particularly opportune time to publish *Beyond Pluto*. John Davies' book on the history and scientific relevance of these denizens of the outer Solar System is aimed at the general reader as well as the astronomer.

The history of the Kuiper belt is both old and recent. As early as the 1930s, after Pluto was discovered, there was speculation about the possible existence of a population of small bodies in the outer Solar System of which Pluto was just the tip of the iceberg. Kenneth

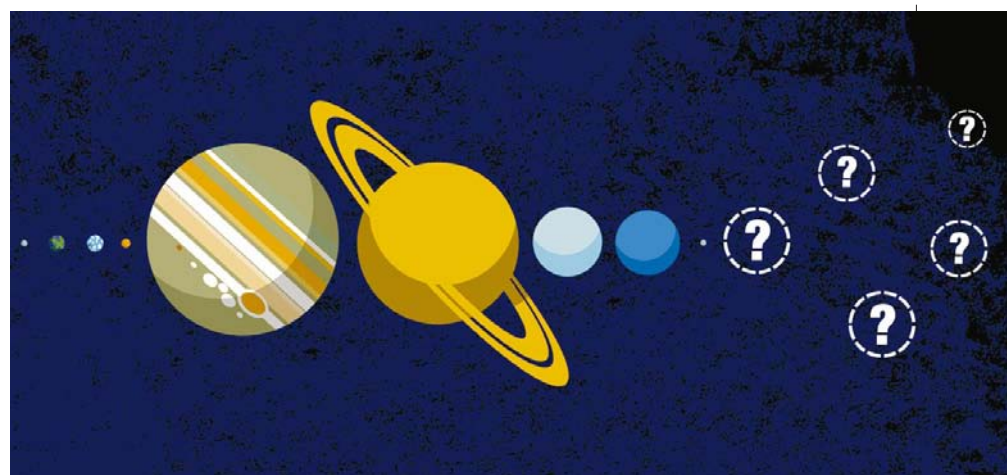
Edgeworth and Gerard Kuiper sketched out these ideas in somewhat more detail in the 1940s, and disagreement persists as to who should get the credit (and the eponymous immortality). To the dynamicist, the history of the Kuiper belt really begins in the 1980s, when models were developed to try to explain the origin and observed distribution of short-period comets (comets that take 200 years or less to orbit the Sun).

To the Solar System observer, depending on whether they are a heretic or a purist, the beginning of our foray into the Kuiper belt would be either the discovery of Pluto in 1930, the discovery of the centaur (outer-planet-crossing asteroid) Chiron in 1977, or the discovery of the first 'classical' Kuiper-belt object in 1992. The Kuiper belt is many things to many people: to the planetary scientist it holds clues to the origin of the Solar System, frozen into its distribution and dynamics during its formation. To the stellar astronomer it is a local example of extended structures we see around some other stars.

All this raises the question of how one can write a history of a topic that is a moving target, and that is, in many ways, still in the first blush of youth, maturing as one writes. Even since this book went to press there have been exciting new discoveries: trans-neptunian objects (TNOs) that are comparable in size to Pluto's moon and the asteroid Ceres; the first binary TNO; TNOs bright enough to be found on pre-discovery plates dating back many decades; and objects on unexpected orbits that have yet to be explained.

John Davies solves this predicament very well by setting out to tell a story and to show the reader that this is an ongoing adventure. The book's dust-jacket describes it as "the fascinating story of how theoretical physicists decided that there must be a population of unknown bodies beyond Neptune and how a small band of astronomers set out to find them". Davies points out that the book is not intended as a textbook, although he hopes (and succeeds, it seems to me) that it gives a feeling for how astronomy is actually done. With that intention, the reader is brought into the story almost as a participant, getting to know some of the personalities involved.

Although the popular idea is that individual personality does not have a role in the impersonal pursuit of science, this story nicely illustrates that it is exactly those unique traits that drive someone to search doggedly over many years for objects they don't even know exist, or to write and wrestle with numerical simulations that must run for months at a time to step through possible billion-year histories of model solar systems. In this book, Davies blends personal histories with scientific theory to give us an astronomical primer on the outer Solar System. He includes a fair amount of self-aware humour about the public and research community's image of astronomy: "Some physi-



cists joke that the fundamental equation of astronomy is that 1 is approximately equal to 10." This personal touch extends to an appendix listing the main belt asteroids that have been named after Kuiper-belt researchers, along with the associated citations describing their work.

Because of the informal conversational style, the story sometimes hops around a bit and reiterates certain themes. Early on, for example, Davies discusses how Edgeworth was the first to put forward the possibility of a disk in the outer region of the Solar System containing small bodies orbiting the Sun. But in the final chapter, more suspects are named (Fred Whipple, Armin Leuschner and Frederick Leonard) who, it could be argued, have the earlier or stronger claim to postulating the existence of small bodies in the trans-neptunian region. If Leonard had got the glory and his term "ultra-neptunian objects" had stuck, then we might now be talking about "UNOs in the Leonard belt".

The introduction of a new concept in the text often leads to tangents that one may find informative, amusing or distracting according to taste (star formation, photometry, the effects of oxygen-deprived observing on Mauna Kea). When discussing the introduction of charge-coupled devices (CCDs) into astronomy as the detector of choice (nearly universally replacing the old standby photographic plate), Davies takes a small sidetrack to explain how they work, which includes an analogy of how a farmer would measure rainfall in his field using labourers with buckets. I found these tangents usually enjoyable, enhancing the informal 'telling a story' style and readability of the book.

Just as the Kuiper belt means different things to different people, there is something here for everyone. For the non-professional, Davies discusses the tools and art of astronomical research in a practical manner. For this level of general consumption, a glossary would have been useful, and the more abstract parts of dynamics, orbital interactions and resonance sweeping, as well as the various orbit diagrams and orbital element

plots may be somewhat difficult for non-experts to disentangle and appreciate. For the expert and research astronomer, this account provides context for the ongoing work in the field as well as an excellent overview of the driving questions and current state of knowledge. For this level, a bibliography would have been beneficial.

But it is a pleasure to have a book that is readable at so many levels and able to describe the concepts and relevance of such a new field of research. With the discovery of the Kuiper belt, our view of our Solar System has expanded in many dimensions at an amazing pace. In the nine years since the discovery of the first TNO (not counting Pluto) to the publication of this book, we have discovered as many objects in the Kuiper belt as it took 100 years to discover in the main asteroid belt. Let's hope that this momentum will continue, with more public recognition of the new worlds to be explored and understood. ■

Joel Wm. Parker is at the Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80302, USA.

The other man to discover evolution

Alfred Russel Wallace: A Life

by Peter Raby

Chatto & Windus/Princeton University Press: 2001. 352 pp. £20, \$26.95

Jane R. Camerini

One can only welcome a new biography of one of Britain's most interesting and least celebrated nineteenth-century naturalists, Alfred Russel Wallace. Known during the twentieth century primarily as the man who spurred Charles Darwin to publish his work on the origin of species, and to biologists as one of the founders of modern biogeography, Wallace has, by and large, been relegated to the periphery of the present-day picture of

Victorian science and culture.

Yet he was the co-discoverer of one of the most significant scientific theories of all time, and a passionate and articulate public intellectual who held forth on an enormous range of political, social and theological, as well as scientific, issues. An advocate of evolution by natural selection, certainly, and of socialism, spiritualism, anti-vaccination and women's rights, Wallace had an unusual following, more chequered and less strictly scientific than that of his more famous peers, such as Thomas Huxley and Darwin.

The anti-establishment leanings of his varied commitments, along with his unease in the social milieu of elite scientific society, contributed to the fact that Wallace never held a regular paid position, and to the fact that he is neglected and even disdained by some historians and scientists. Although celebrated writers such as A. S. Byatt and Andrea Barrett have found inspiration for their fiction in Wallace's intriguing life, he has not until now received his due from historians. Peter Raby's energetic account of Victorian scientific travellers, *Bright Paradise* (Princeton University Press, 1996), put Wallace's journeys to tropical Brazil and Indonesia in their historical and cultural context. His new book provides a focused and balanced narrative of Wallace's life, in which his exotic travels are placed in the context of his eventful life story.

Born in 1823 in Wales, Wallace made his way from land-surveying in early industrial Britain to working as a naturalist in the wilds of the Amazon basin by the age of 25. A self-taught natural scientist, paying his way by collecting insects and birds to sell to the then-thriving trade in specimens of natural history, Wallace's immodest goal was to find a theory to explain the origin of species.

Raby skilfully draws readers into Wallace's story, which ranges from an unremarkable upbringing to the rapids of the Rio Negro, from tropical shipwreck to strolling in London's Regent's Park. We follow Wallace as he braves London's scientific social world, and departs once again, this time to the islands of the Malaysian archipelago in search of new evidence for his growing "obsession", as Raby calls it, with understanding the laws that govern the variation and distribution of species.

Raby relates with a biographer's remove Wallace's discovery of the dynamic law — subsequently called natural selection — that underlies the formation of new species. And he tackles Wallace's return to England and his subsequent infatuation with spiritualism using a combination of insight and authorial scepticism. Of note is Raby's perception that, when investigating seances and other spiritualist phenomena, "Wallace's conviction threshold was lowered whenever he came across some apparent fact or reference involving a member of his family ... For Fanny

[Wallace's sister] and Alfred, the certainty that they were in touch with their dead brothers and sisters formed the bedrock of their spiritualist convictions." Given that six of Wallace's eight siblings were dead, this adds a personal dimension to Wallace's belief. Subsequent developments of Wallace's scientific and social commitments, and his journey to North America, are also covered.

Raby expands our understanding by his thorough research of manuscript material, particularly letters and other documents in the Wallace family archives, as well as a conscientious use of published and archived letters of his contemporaries. His familiarity with Victorian literature and drama adds sensitivity to his portrait of Wallace, who read widely throughout his life in spite of his lack of higher formal education.

Well researched and ably written — and a definite improvement on the few existing biographies of Wallace — the book is nonetheless mildly disappointing, particularly for historians of science. Raby's thick-brush treatment of the science of Wallace's day is superficially lively, providing an overview from the outside, but only begins to relate the enormous upheaval surrounding the new findings and new interpretations of the history of humankind that so dominated the public intellectual scene in the 1860s. Here, as elsewhere, Raby seems to skim over recent work in the history of science.

In recent decades, the bar has been raised for scientific biography with Adrian Desmond and James Moore's *Darwin*, Janet Browne's *Charles Darwin* and Adrian Desmond's *Huxley*, in which biographical narratives are embedded in finely textured political and social contexts. Raby's strategy of sticking fairly closely to the rich set of sources he has gathered proves to be a double-edged sword. He provides readers with a well-documented biography that covers the principal events of Wallace's life, but a deeper and broader historical sensibility — in which connections are made among the activities of scientists, the content of scientific work, and the broader social, cultural, political and scientific contexts — is wanting.

I would recommend this tastefully illustrated book to general readers as a good introduction to one of Britain's more charismatic and difficult figures of the nineteenth and early twentieth centuries. Historians of science will probably find some new information and will continue to wonder about the unresolved questions of Wallace's role in the history of science. ■

Jane R. Camerini is in the Department of the History of Science, University of Wisconsin, Madison, Wisconsin 53706, USA. Her forthcoming book The Alfred Russel Wallace Reader: A Selection of Writings From the Field (Johns Hopkins University Press) will be published later this year.



Unlocking nature's ancient secrets

The Molecule Hunt: Archaeology and the Search for Ancient DNA

by Martin Jones

Allen Lane: 2001. 280 pp. £18.99

Tomas Lindahl

Traditional archaeological research describing skulls and pottery fragments has expanded in recent years to the use of genetic and biochemical methods in pursuit of the origins of humans, domestic animals and agriculture. The rewards have been great, and structural determinations of DNA and other macromolecules in ancient materials have transformed the field almost as much as DNA analysis has enhanced forensic medicine. For example, an unknown genius in southeastern Turkey, about 10,000 years ago, must have dramatically altered human life patterns by domesticating wild cereals, thereby initiating agriculture. The later exploitation of wild maize in America may have been a more gradual process.

DNA sequencing has enabled us to trace the origins of horses, dogs and cattle as human domestic animals in a detail that was hard to imagine a few years ago. The huge and fierce aurochs, depicted in cave paintings and at ancient grave sites, now emerges as a kind of bovine Neanderthal, extinct and not a direct precursor of modern cattle.

Martin Jones tells the intriguing story of the new field of bioarchaeology in this timely book. From his position as the George Pitt-Rivers Professor of Archaeological Science at Cambridge University, and chairman of the UK Ancient Biomolecules Initiative, he has had a broad and unique overview of the rapid evolution of this new area of research,

and his lucid and authoritative writing conveys the excitement of the field. Human migration patterns in ancient times, the causes of terrible epidemics in the distant past, and many similar questions can now be defined by DNA analysis.

Because DNA decomposes by chemical decay with time, it has not been possible to extend molecular analysis into the very distant past, beyond about 100,000 years. However, the development of ultrasensitive sequencing methods for proteins, which are more stable, might push the time barrier a little further. The oldest DNA fragments yet recovered are short stretches of mitochondrial DNA from woolly mammoths and mastodons. Not surprisingly, their DNA sequences show great similarities to modern elephants. But the interesting fact has emerged that there were many quite distinct types of mammoths, and so elephants represent merely a small remnant of a once-great pattern of diversity.

The retrieval of DNA sequences from Neanderthal bones, slightly less old than the mammoths, has been a particular triumph in this field of research, and has greatly elucidated the distinct differences between early humans and Neanderthals. A decade ago, before it was realized that DNA cannot be preserved for millions of years and contamination with modern DNA is a major technical problem, there were several overly optimistic reports on the apparent recovery of DNA from insects entombed in amber and other potential sources more than 100 million years old.

The recovery of DNA from a dinosaur bone made the headlines in both the scientific and daily press. It was followed by the dampening discovery that this particular DNA sequence was identical with a piece of human junk DNA, a fragment of mitochondrial DNA that had found its way to the cell nucleus and been integrated as a pseudogene during evolution. It is not often that molecu-

lar biology turns into farce, but in the published discussion that attempted to explain this amazing finding, the possibilities were raised that a roving dinosaur had entered the laboratory and contaminated a piece of equipment, or that in the distant past intercourse had taken place between a dinosaur and a human individual. Jones recounts the absurd story totally straight.

The contamination problem remains the greatest difficulty in this dynamic new area of research. In a recent conversation with one of the leaders of research into ancient DNA, I was told how he had handed a famous palaeontologist a brownish, ancient human skull from a museum collection with a query as to whether the skull had been coated with varnish. The palaeontologist gripped the skull with both hands, stretched out his tongue, licked the top of the skull and decided from the taste that the skull indeed had been varnished. In the process, large amounts of modern human DNA would have been transferred to the skull, hampering any future search for traces of remaining ancient DNA fragments. Perhaps archaeologists should change their field techniques and use sterilized instruments, protective clothing, plastic gloves and face masks during their digs.

The book's final chapters describe how the molecule hunt continues. Jones is more circumspect and less authoritative here dealing with microbiological problems, in particular with regard to recent bizarre claims that very ancient bacterial spores might be revived. It requires considerable wishful thinking to believe that bacteria recovered from an excavated fossil or salt crystal are as old as the fossil itself, especially if the bacteria are common extant ones. The only credible way to retrieve very ancient inherited information would be by identifying a slowly propagating organism, halted in its evolution and possessing unique properties clearly not shared by other existing species.

Perhaps some cold and hostile ecological niche, with a reducing, anoxic environment similar to that of Earth a couple of billion years ago, continues to shelter a lurking primitive microorganism that has retained an inefficient early version of the genetic code, or even a chemical precursor form of modern DNA. The isolation and propagation of such a molecular coelacanth would be received with astonishment even by the scientific world, and would teach us much about the origins of life.

The chances of finding something may not be great, but very little work has been done on anaerobic systems from this point of view. In fact, the odds seem rather better than those on the futile and costly search for life on Mars.

Tomas Lindahl is at the Clare Hall Laboratories, Imperial Cancer Research Fund, South Mimms EN6 3LD, UK.

Modern values in post-modernism

The One Culture?

A Conversation about Science

edited by Jay A. Labinger & Harry Collins

Chicago University Press: 2001. 322 pp.

\$17, £12 (pbk)

John Ziman

Without a doubt, people ought to have a better understanding of the status of scientific knowledge. Yet most scientists know little — and care even less — about the 'science wars'. This courteous exchange of views between a few of the active partisans may help them to resolve some of their mutual misconceptions, but does little to enlighten the ignorant eavesdropper. What is all the argy-bargy about?

The bogey word, and one that is not much mentioned in *The One Culture*, is 'post-modernism'. Despite all the intellectual atrocities committed in its name, this was a legitimate reaction against the doctrine of 'modernism'. It questioned the idea that formal rationality was the sure way to happiness, goodness and beauty. At first, the 'culture wars' raging in the arts and humanities seemed of no concern to natural scientists, but the post-modernists also unmasked many social pathologies that claimed 'scientific' authority. Science itself, the traditional stronghold of modernism, became a strategic target of their critique.

The gates of the citadel were inadvertently opened by Thomas Kuhn in 1962. His admirable historical study of scientific revolutions could be read as a scholarly justification for complete metaphysical relativism. He seemed to be saying that apparently secure scientific knowledge was never more than the collective opinion of a particular social group, and could thus be debunked by reference to the views of other groups. Philosophically speaking (as several of the contributors to *The One Culture* point out), total scepticism is old hat, and essentially vacuous. Nevertheless, it remains the principal (unsmart) weapon in the 'science wars', and is not adequately countered by equally naive mantras (also uttered here) that scientific knowledge is really, really true.

More constructively, Kuhn was drawing attention to the social factors that influence the production of scientific knowledge. Strangely, this influence had been systematically denied by philosophers. Perhaps they were so focused on the hypothetical ideal of finally revealing scientific truth that they felt they could overlook the messy human processes along the way. Unfortunately, scientific knowledge is always on the road, so its current state is inevitably affected by where it is or has just come from — that is, by the ideas and interpersonal communications of



the researchers and research institutions that produce it.

Most participants in the debate agree that science has nothing to fear from a closer study of its social features. Indeed, they accept that it is only by understanding the creative and critical practices by which scientists cooperate and compete that we can fully appreciate the status of scientific knowledge — the extraordinary reliability of some parts of it and the horrible uncertainty of much of the rest. How come, then, that sociology is widely perceived as 'anti-science'?

This conversation doesn't resolve the puzzle. In the book, the so-called 'sociology of scientific knowledge', or SSK, fills the stage. But despite their self-proclaimed relativism, its proponents repeatedly insist that they too are faithful supporters of the scientific enterprise. In any case, the innocent reader should not accept their claim to be speaking for the whole field of 'science studies'. As the admirably broad-minded contributions of Michael Lynch and Steven Shapin show, SSK is actually only one of the many guerilla bands that operate in this wide-open academic territory.

What the SSKers say again and again in this rambling conversation is that their relativism is strictly "methodological". As their avowed leader Harry Collins puts it, they are carrying out an "amoral analysis", in which they "leave the scientists to decide on the truth" while "recording the argument without taking sides". The practices and beliefs associated with a research project in physics should be treated "symmetrically" with those associated with, say, the casting of a horoscope or the detection of a witch among the Azande of Sudan. Indeed, the adherents of SSK say that they are just ethnographers, like Edward Evans-Pritchard, who adopted a similar stance in order to demonstrate the innate rationality of pre-scientific cultural systems.

Unfortunately, as most social scientists

now acknowledge, this highly desirable objectivity is nowhere near attainable. The conversation does not include any sociologists or ethnographers outside the narrow speciality of the sociobiology of science, so it is never pointed out that outsiders don't necessarily have a clearer view of a culture than the people who live inside it. In practice, methodological relativism shades into rhetorical affectation, dangerously compromising the sincerity of the 'science studier' in the eyes of the 'studied scientist'. Thus, to preserve a semblance of consistency, SSK overbalances into full metaphysical relativism, where its pro-science protestations sound very thin.

What has SSK actually taught us? Principally, it has shown us that scientific research is 'really' just like any other human activity. Big deal! In fact, this unsurprising outcome was written into the initial script. SSK aims to tell everyone what they should know about science, but aborts its own mission by perversely bracketing out the social peculiarities that differentiate science from other modes of knowledge production. Thus, it has nothing to say about the practices that seem to make scientific knowledge so remarkably reliable for certain purposes.

SSK is just too conceptually limited to answer the questioning title of this muddled, muddling book. The 'pro-science' conversationalists here are all physical scientists, so they fail to see that the natural world has many different aspects, each inspiring a somewhat different scientific culture. Or should one say that the natural and human sciences, including SSK itself, are all subcultures of modernism, all challenged by the post-modern critique, and all really on the same side in the science wars?

John Ziman is at 27 Little London Green, Oakley, Aylesbury HP18 9QL, UK. His most recent book is *Real Science: What It Is and What It Means* (Cambridge University Press).

A life of good taste

Fine Wines and Fish Oils: The Life of Hugh Macdonald Sinclair

by Jeannette Ewin

Oxford University Press: 2001. 240 pp. £25

Tom Sanders

The nutritionist Hugh Sinclair (1910–90) was the archetypal Oxford don: erudite, urbane but disorganized, and an accomplished and witty speaker with an astonishing repertoire of nutritional anecdotes. This biography, however, which borders on being a hagiography, focuses on Sinclair's work, particularly his Oxford Nutrition Survey, his role in stimulating research on essential fatty acids and his unsuccessful struggle to establish a department of nutrition at Oxford University.

Sinclair's father was an elderly army colonel, who died when Sinclair was at boarding-school, and his mother, a member of the wealthy Scottish aristocracy, lived with him until her death in 1969. Sinclair was proud of his ancestral heritage and was an inveterate snob. Educated at Winchester and Oxford, he cultivated a taste for entertaining and fine wines while at Magdalen College. Having purchased life membership of the Middlesex Cricket Club aged nineteen, and inveigled himself into membership of the Athenaeum Club, Sinclair had a sound base from which to network with the influential scientists and magnates of his day. He read widely, travelled extensively and was able to meet some of the world's great nutritional scientists. Indeed, it could be said that he spent more time travelling and attending conferences than doing research.

Sinclair was director of the Oxford Nutrition Survey, which aimed to assess the prevalence of malnutrition among the British population. As it evolved, the survey grew more complex and became embellished by numerous studies of dietary interventions, generating enormous amounts of social, clinical and biochemical data. In 1943, during a critical phase of the survey, Sinclair absented himself for three months to attend scientific meetings and visit colleagues in the United States and Canada. He failed to work out how the accumulating data could be analysed, and the survey remained unanalysed and unpublished in his lifetime.

In 1945, Sinclair was given the rank of brigadier and sent to Holland and Germany to work as part of a team assessing the nutritional status of the Dutch and German populations. For this work he received the Medal of Freedom with Silver Palm from the United States.

While touring Canada in 1943, Sinclair met Group-Captain Tisdale of the Royal Canadian Air Force, who invited him the



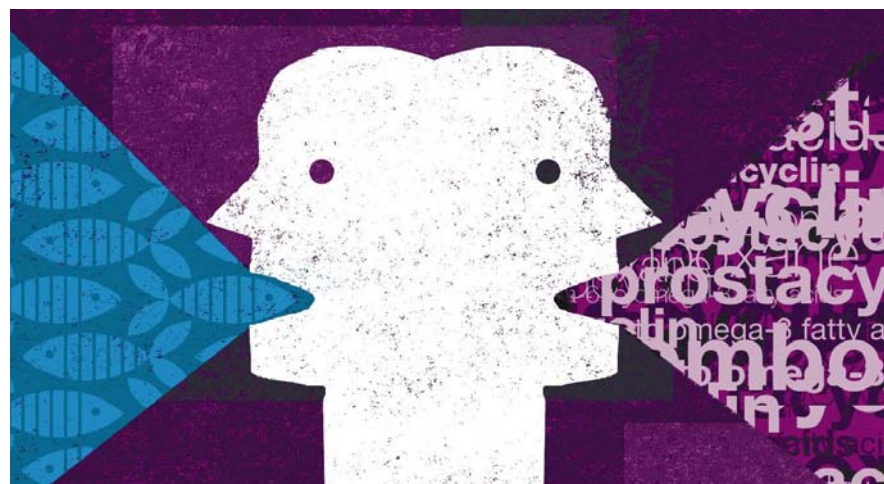
following year on a short expedition to observe whether riboflavin deficiency was related to snow blindness among the Canadian Inuit. It is reported that he failed to keep any written record in his diaries but embellished the tale in later life to suggest that he had joined the expedition because he was interested in the fact that the Inuit diet was high in fat, rich in essential fatty acids, and yet the Inuit were free from heart disease. During this period of his life, Sinclair's work was concerned with thiamine and diseases of the nervous system, and there was no evidence of his having any interest in cardiovascular disease and dietary fat. His epic letter to the *Lancet* in 1956, in which he suggested that cardiovascular disease was caused by a deficiency of essential fatty acids, was an important stimulus to future research.

But Sinclair was blinkered by the deficiency paradigm. And the book perpetuates the myth that he was responsible for drawing attention to the cardioprotective properties of omega-3 fatty acids. What he failed to note was that the balance of omega-6 to omega-3 fatty acids was important to health. Indeed, for many years he promoted the consumption of a diet high in omega-6 fatty acids. The major impetus for cardiovascular research on the omega-3 fatty acids arose from the work of Salvador Moncada and John Vane on prostacyclin, and Philip Needleman on thromboxane, published in 1976, three years before Sinclair embarked on his Eskimo diet to demonstrate the effects of omega-3 fatty acids from fish oils on haemostasis.

Sinclair's life-long ambition was to establish a department of nutrition at Oxford. He was appointed reader in human nutrition there in 1951. There is little doubt that he was an able scholar, but his ability as a research scientist is questionable because of his lack of attention to detail and failure to publish his results in peer-reviewed journals. He was a prolific letter-writer and collector of manuscripts (including a collection of erotica) and, following his death, these sold for more than £85,000 (US\$124,000).

Like a few other famous nutritionists, such as Boyd Orr and Robert McCarrison, Sinclair liked to dabble in the politics of food and influence national policy. But his outpourings tended to be based on belief and theory rather than evidence and he was openly contemptuous of the work of his contemporaries, such as John Yudkin and Elsie Widdowson. But to his credit, Sinclair truly understood the complexity of the relationship between diet and health and recognized the need for a multidisciplinary approach.

As a scientist, he came to be regarded as a dilettante; his research lacked focus and was unsystematic. This, coupled with his failure to complete projects and produce peer-reviewed publications, and his sniping at influential contemporaries, eventually resulted in his ejection from Oxford's Department of Biochemistry in 1956 by Sir Hans Krebs, and his



readership was not renewed in 1958. For the rest of his working life, Sinclair remained in the wilderness of his self-styled National Institute of Nutrition, which was situated in the grounds of his home at Lady Place in Sutton Courtenay. On his death, he bequeathed his estate to establish a chair in nutrition at Oxford which the university declined. The offer was eventually taken up by the University of Reading, where the Hugh Sinclair Nutrition Unit thrives under Christine Williams.

This is no detective story: there are no elegantly designed experiments or startling discoveries. It is a salutary warning to nutritionists that scientific progress is made by good experimental design and meticulous attention to detail and not by travelling the world on lecture tours.

Tom Sanders is in the Department of Nutrition & Dietetics, King's College London, Franklin-Wilkins Building, Stamford Street, London SE1 9NN, UK.

In for the count

Mathematical Mountaintops: The Five Most Famous Problems of All Time
by John Casti

Oxford University Press: 2001. 288 pp.
£19.95, \$25

Simon Singh

The recent boom in mathematics bestsellers has contributed a great deal towards raising the public profile of the subject. But such books ignore a significant section of potential readers, namely those who have more of a mathematical background than the general reader but who are not professional mathematicians. Such mathematical enthusiasts have no doubt enjoyed some of the popular books, but would really prefer a more technical treatment. This is exactly what John Casti provides in *Mathematical Mountaintops*. It is neither a textbook nor a pop maths book, rather it is a serious in-depth look at the great problems of mathematics.

Casti has picked "the five most famous problems of all time", and spends 30 to 40 pages describing each one. The problems are Hilbert's tenth problem, the four-colour problem, the continuum hypothesis, the Kepler conjecture and Fermat's last theorem. Each of these has now been solved, so, in addition to outlining the problem, the author is able to explain the solution and recount the story behind it. Four of the problems have been written about extensively elsewhere, but perhaps not with Casti's balance of technical explanation and background narrative.

Casti's remaining problem, the Kepler conjecture, has (to my knowledge) not been written about since the recent announcement that it has been proved, and provides perhaps the most interesting chapter. The problem dates back to 1606, when Johannes Kepler posed a question in a paper for his patron Johann Matthäus Wacker of Wackenfels, Knight Bachelor. Kepler asked, what is the most efficient way to stack spheres so as to minimize the spaces between them? Alternatively, what is the best way to pack oranges in an infinite box? Kepler proposed that the best arrangement was the face-centred cubic lattice, in which every sphere in the first layer is surrounded by six others, and each subsequent layer is built by putting spheres in the dimples of the layer below. This arrangement has a packing efficiency of 74.048%. Grocers, who traditionally stack oranges in this way, suspected that Kepler was right, but it took mathematicians almost four centuries to prove it.

There were some notable milestones along the way. In 1694, Isaac Newton and the Scottish astronomer James Gregory argued about the sphere-kissing problem: what is the maximum number of spheres you can place simultaneously in contact with a central sphere? Newton said that the answer was twelve, which is easily achievable, but Gregory was convinced that it was possible to squeeze in a thirteenth sphere. Newton turned out to be right, but this took 180 years to prove.

The Kepler conjecture was eventually proved in 1998 by Thomas Hales of the



University of Michigan, following an approach developed in the 1950s by the Hungarian mathematician Lázlo Fejes Tóth. Hales showed that it was possible to determine the maximum packing efficiency by analysing a cluster of just 50 spheres. Each sphere has a position in three-dimensional space, so the packing efficiency depends on an equation containing 150 variables. Maximize the equation and you have the maximum packing efficiency, although this is not a trivial problem.

Hales and his graduate student Samuel Ferguson maximized the equation, thanks to some clever mathematical short-cuts and a tremendous computing effort which used a program that relied on three gigabytes of storage. The fact that Hales's proof relies so much on a computer gives rise to one of the most interesting aspects of Casti's book, namely the validity of computer proof.

In fact, although each chapter is about a particular problem, these problems are used to convey broader ideas about the nature of mathematics in general, its motivation and objectives, its culture and rules. Over the past 25 years, the use of computers has changed the nature of mathematics, solving some previously intractable problems, but sowing discord within the mathematical community.

For example, Casti writes about the four-colour problem, which was also solved with the aid of a computer. After Kenneth Appel and Wolfgang Haken proved it in 1976, their lectures were sometimes met with hostility from their colleagues and some professors barred their graduate students from talking to the notorious duo because this was a computer proof, not a traditional proof. A rumour began to spread that there was a bug in the program, but no bug was ever found. In fact, it was the hand-generated part of the proof that contained errors, none of which turned out to be serious.

Casti explains that a good proof should satisfy three criteria. First, it should be convincing — in other words, mathematicians should believe it when they see it. Second, it should be formalizable, which means it can

be incorporated within an established logical framework. Third, it should be surveyable, or capable of being understood, studied, communicated and verified by rational analysis. However, a computer proof fails to satisfy the third requirement, at least in any traditional sense. In the past, mathematicians could work through a proof line by line and explain it to one another. In a computer proof, the broad approach can be checked, but the detailed calculations are embedded within computer code and can be performed only by a microprocessor. The proof, to some extent, has to be taken on trust.

For readers who are particularly inspired by *Mathematical Mountaintops* and who have some spare time, Casti's final chapter briefly discusses the unsolved Clay Institute problems. Last year, the Clay Mathematics Institute held a press conference and identified seven problems that were crucial to mathematics in the new millennium. This resonates with German mathematician David Hilbert's list of outstanding problems announced in 1900. Whoever solves any of the Clay problems will win a prize of \$1 million. But more importantly, they will earn a place in mathematical history and perhaps their own chapter in a subsequent edition of *Mathematical Mountaintops*. ■

Simon Singh is a science journalist and broadcaster based in London. He is author of The Code Book: The Secret History of Codes & Code-Breaking (Fourth Estate) and Fermat's Last Theorem: The Quest to Solve the World's Greatest Mathematical Problem (Fourth Estate).

Is anyone out there listening?

The World According to Pimm: A Scientist Audits the Earth
by Stuart Pimm
McGraw-Hill: 2001. 304 pp. £18.99, \$24.95
Harold Mooney

Those who study global change are well versed in the sobering statistics of the enormous impact of humans on the Earth — the dramatic change in the chemistry of the atmosphere, the massive alteration of the surface of the land, the diversion and despoiling of a large fraction of the available fresh water, the depletion of ocean fisheries, the homogenization of the Earth's biota and the extirpation of large numbers of species. These scientists share a sense of frustration, however, about the fact that the general public and policy-makers are not grasping the significance of these changes nor acting to alter these trajectories for the well-being of future generations. In *The World According to Pimm*, Stuart Pimm is attempting to enlarge the army of scientists working on

these issues and to engage the public in the dialogue about what is happening to the biotic resources of the Earth and how we need to change our trajectory of development in order to build a sustainable world.

To accomplish his task, Pimm has written an engaging and important book. He might well have called it "The World According to Peter, Paul and Pauly as viewed by Pimm", as the bulk of the book is devoted to an analysis of studies involving Peter Vitousek, Paul Ehrlich and Daniel Pauly. The book takes on a bold challenge — an accounting of the productive capacity of both the land and of the oceans and how humans have modified it. Pimm then documents the extent to which humans are utilizing the available fresh water and, finally, the impact of humans on the biodiversity of the Earth. He uses very simple mathematics and language to drive home the point that humans have very substantially modified the Earth's biotic resources.

The first part of the book is based on the dramatic retelling of the story first published by Vitousek and colleagues in 1986 in *BioScience*, in which they calculated that humans already use 40% of the primary productivity of the land. That is a sobering figure in view of the projections of growth of the human population during this century. There was some difficulty in getting this penetrating study published, but once it was, it was used widely. For quite a while it was impossible to go to a meeting related to the environment where the statistics of this paper were not used in the opening address. Subsequent attempts to check and recalculate the numbers, as Pimm has done, have borne out the original thesis. Pimm's recalculations are engaging; he does a very good job of making sure that the general reader can clearly understand the basis for these estimates of human use of the biosphere.

The second part of the book deals with human use of the Earth's water resources and builds on work by Ehrlich and his colleagues Sandra Postel and Gretchen Daily. They have shown that humans are using a sixth of the total estimated runoff (50% of that available), and in doing so have drastically altered the Earth's rivers — its plumbing system — especially in the Northern Hemisphere.

The work of Vitousek and colleagues inspired Pauly and a colleague to do a similar



study of the human use of the oceans. They estimated that a third of the oceans' production goes to support human fisheries — another stunning number in view of the fact that, not too long ago, the ocean fisheries were considered to be inexhaustible. Large numbers of fisheries are now depleted or endangered. We are clearly at the limits of ocean exploitation. Further, as Pauly and others have shown, we are now "fishing down the food chain", as the large predatory fish have been depleted.

Pimm concludes his book with a sobering analysis of species losses, his speciality. He ends his analysis with a plea for the preservation of those areas of the world that are particularly rich in species, the so-called 'hot spots'. He claims that we do have the means at hand to "defy nature's end" by purchasing and protecting examples of these areas.

In recent years the international scientific community has made an enormous effort to carefully document, through the International Panel on Climate Change, the impact of humans on the Earth's climatic system. The panel has concluded that human action is indeed altering the climate of the Earth with profound consequences for the way we will live our lives. Pimm does not deal with the consequences of these changes, nor with the escalating disruptions due to the biological homogenization of the Earth's biota. He purposely omitted considering future changes and concentrated on what we know now and for certain. Putting where we are today together with what the future most probably holds for the biotic systems on which we depend certainly gives cause for concern. Let's hope that Pimm's book will educate more of us on what the stakes are and why we need to move urgently towards a more sustainable use of our resources than we are seeing today. ■

Harold Mooney is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

Discourse and discord

Reconciling Science and Religion: The Debate in Early Twentieth Century Britain

Peter J. Bowler

University of Chicago Press: 2001. 496 pp. \$40, £24

Rebuilding the Matrix: Science and Faith in the 21st Century

Denis Alexander

Lion Publishing: 2001. 544 pp. £20

Geoffrey Cantor

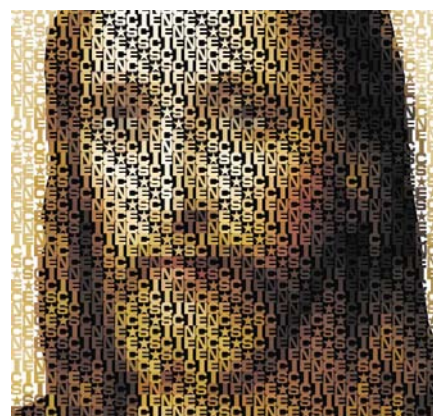
These timely books form part of the rapidly expanding body of literature on the interrelations between science and religion.

Bowler, an historian of science who disavows any religious convictions, has addressed a significant lacuna in existing scholarship. Whereas much has been written on the Victorian controversies, often focusing on the scientific naturalism and materialism spearheaded by Thomas Henry Huxley and John Tyndall, little attention has been paid to the opening decades of the twentieth century. The Scopes trial of 1925, which considered whether the state of Tennessee could prosecute John Scopes for teaching the theory of evolution in a public-school science class, is the best-known event of this period. But it had little impact on British writers, who inherited a very different set of religious and social assumptions compared with their American counterparts.

Bowler's main contention is that, during the early twentieth century, discussion of science-religion issues in Britain was dominated by reactions against Victorian naturalism. With agnosticism and materialism considered as *passé*, many scientists, theologians and popular writers sought a new *rapprochement* between science and religion. Although a few atheists, such as E. Ray Lankester and Arthur Keith, continued the assault on Christianity, they were considerably outnumbered by those who advocated some form of synthesis.

Yet there was no consensus on how this synthesis should be achieved. Responses ranged from Oliver Lodge's advocacy of spiritualism, which linked the worlds of matter and spirit, to Arthur Eddington's enthusiastic embrace of the new physics in 1927 which, he claimed, only then made religion "possible for a reasonable scientific man". Whereas Bishop Ernest Barnes advocated an effervescent mixing of science with progressivist, evangelical Christianity, churchmen of a more conservative stripe, such as Charles Gore, sought a reconciliation that preserved traditional Christian doctrines.

One of the great strengths of Bowler's book is that it demonstrates the richness of science-religion discourse during this period and provides a helpful map of the terrain. Not only does Bowler discuss the more eminent scientists and theologians, whose positions have previously been analysed, but he sheds considerable light on several lesser-known figures. He also shows that issues of science and religion were subject to comment and controversy far beyond both professional communities. The public, for example, encountered these issues in newspapers, in the periodical press and in books aimed at a wide readership. Thus, authors as diverse as H. G. Wells, G. K. Chesterton and Bertrand Russell are included in Bowler's analysis. Given the considerable range of commentators discussed, it is not surprising that Bowler has paid less attention than he might to the broader social and political movements of the period that impinged on



the religious and scientific life of Britain.

The lively interest in issues of science and religion during the opening third of the century and the attempt to transcend the Victorian impasse had given way to a less favourable atmosphere by the late 1930s. Increasing politicization and the rise of a new generation of scientists with different agendas led to a greater polarization of positions and a lowering of interest in the topic. Although the discussion of science and religion received a further boost in the postwar period, largely through Charles Coulson's writings, it has regained prominence only during the past few years. The other book under review reflects this trend.

As a molecular immunologist and committed Christian, Denis Alexander is troubled by two popular prejudices of our age. One is the opposition to science (which has sometimes been promulgated in the name of religion); the other is the assumption that science and religion are locked in necessary conflict (a belief that is often advanced in the name of science). To counteract both prejudices, Alexander has written an introductory and wide-ranging text that is intended to encourage the reader to adopt a more tolerant position, particularly on the issue of science and religion, on which atheists such as Richard Dawkins and Peter Atkins have grabbed the public's attention.

To engage this topic, Alexander has to clear the undergrowth by introducing an extensive range of issues that concern both scientific knowledge and religious knowledge (as opposed to belief). Although he acknowledges the role of social factors, he adopts a critical realist stance towards science, rejecting not only the dogmatic image of science that is often conveyed to the public, but also the threat of relativism. On this and other epistemological issues, he conceives close parallels between science and Christianity. Although some of these correspondences are more convincing than others, they not only refute the much-vaunted conflict thesis but also indicate how science and religion can be located in the same epistemological frame.

Alexander also insists that the theory of

evolution has neither religious nor metaphysical implications, and should therefore not be extrapolated into such areas as evolutionary ethics. Although many practising scientists will support his position, it cannot be implemented because, as Bowler's study shows, reflection on the wider connotations of scientific theories has proved both important and attractive to many writers, including some scientists. Alexander's strategy of guarding science against questionable implications nevertheless enables him to combat both the anti-religious scientism of atheists such as Dawkins and the anti-evolutionism of creationists. Proponents of both positions, he claims, have misunderstood both science and religion. Indeed, proponents of these twin heresies have misinterpreted the opening verses of Genesis, which should not be read as a proto-scientific account describing how the world came into existence, but should be understood in the context of the creation stories that were current at the time it was written.

Alexander also provides a long and informative overview of the history of science-religion interactions from the ancients to the Victorian controversies surrounding darwinism and materialism. From this survey he draws the possibly over-optimistic conclusion that, in the past, science and religion have interacted in many different ways and to the benefit of both parties. Interestingly, his historical survey stops short of the period that is illuminated by Bowler's book. Although some of Alexander's arguments reflect those used in the period discussed by Bowler, Alexander's contribution to science and religion helps to move the subject into the twenty-first century. ■

Geoffrey Cantor is at the School of Philosophy, University of Leeds, Leeds LS2 9JT, UK.

Washing one's hands of controversy

Prometheans in the Lab: Chemistry and the Making of the Modern World

by Sharon Bertsch McGrayne
McGraw-Hill: 2001. 243 pp. \$24.95

Martyn Poliakoff

When Prometheus stole fire from the gods, he had to endure their wrath. But he did not have to worry about the effect of fire on the environment when he released it on Earth. That was the users' problem. Things were different for the revolutionaries of industrial chemistry, the 'Prometheans' of the title. The gods did not interfere but the environmental consequences came back to haunt them.

However, the Prometheans fought back. When Thomas Midgley, the inventor of lead-

ed petrol, was asked at a press conference whether his new additive was dangerous when spilled on the hands, he "called for tetraethyl lead. Pouring it into his palms, he washed his hands with it, and then held the bottle to his nostrils for more than a minute. Drying his hands on his handkerchief, he announced, 'I'm not taking any chance whatever.'" Midgely knew it was untrue but it was the message the public wanted to hear. Lead cured the knocking that crippled car engines and reduced petrol consumption. Lead was needed to drive the United States through the twentieth century.

The balance between consumer benefit and environmental impact is not clear-cut, and this is the crux of Sharon Bertsch McGrayne's ambitious book. Her case is built around the lives of outstanding chemical inventors from the eighteenth century to the present day. Their lives illustrate the interplay of consumer demand, corporate greed and environmental fallout.

The inventors are not all well known but they have been chosen with care, albeit with a slight bias towards the United States. They include: Nicolas LeBlanc, the inventor of an extremely smelly and polluting method for producing washing soda, who lost his patents during the French Revolution; William Perkin, the 18-year-old English inventor of the first synthetic purple dye, who retired a millionaire at the age of 35 to devote himself to scientific research; Norbert Rillieux, an African-American chemical engineer and cousin of the painter Degas, who introduced a revolutionary steam evaporator for refining sugar; Edward Frankland, the British founder of analytical chemistry and champion of clean water, who suffered from the life-long stigma of his illegitimacy; Fritz Haber, the German inventor of the eponymous process for making ammonia from atmospheric nitrogen and chief proponent of chemical warfare in the First World War, whose first wife committed suicide in disgust at his war role (or was it at his infidelity?); Thomas Midgley, the 'lead handwasher', who also invented ozone-depleting CFC refrigerants and arranged to be strangled by the contraption that he devised to turn him in bed after he was struck down by polio; Wallace Carothers, inventor of nylon and founder of polymer science, who was an alumnus of the same little-known college as Carl Djerassi, inventor of the contraceptive pill; Paul Müller, the Swiss inventor of the insecticide DDT, whose boss contested his Nobel prize; and the US geochemist Clair Patterson, who established the age of the Earth and then went on to show that it was polluted even at its utmost ends by lead from gasoline.

Bertsch McGrayne is a science journalist and really comes into her own when the book moves into the twentieth century. She has missed some well-known anecdotes that could have enlivened the earlier chapters,

such as Perkin's luxuriant beard from which debris was reputed to fall into his solutions to nucleate crystallization. But many fascinating stories are there, including Haber's vain and highly secretive attempt to fund Germany's war reparations by extracting gold from sea water.

One of her main arguments, which I strongly support, is that we should not apply environmental judgements to historical situations without considering their humanitarian context. Thus, the first large-scale application of DDT was made during a typhus epidemic in Naples in December 1943. Within three weeks, the civilian death toll had halved, the first time that a typhus outbreak had been controlled in winter. Similarly, one of the main benefits of CFCs was safe refrigeration, which not only reduced food poisoning but also allowed the widespread use of vaccines, for example in eradicating smallpox. That said, the book contains several examples in which the needs of corporate profit and human welfare clearly diverge.

Bertsch McGrayne explains how, as a Jew, Haber was airbrushed out of history by the Nazis. However, his former colleagues circumvented the Nazi boycott of his memorial service by sending their wives in their place. But she fails to point out the interesting moral contradiction in Haber's work. His ammonia process was invented to make fertilizers to relieve hunger, yet it was used for explosives that killed millions. He invented mustard gas as a weapon, but its derivatives have saved the lives of more cancer patients than it ever killed soldiers on the battlefield.

This is not a chemistry book; the chemistry it does contain is painfully spelled out. It is well researched, but the author eschews footnotes and her sources are hidden in a somewhat idiosyncratic "annotated bibliography". There are a few line diagrams but no pictures, which is a pity in a book that is so centred on personalities. From the prelude to the quaintly named "postlude", it is a compelling read, provided that you persist beyond the opening chapters. ■

Martyn Poliakoff is at the School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK.



Bridge or ravine?

Ideas that cross the border between scientists and non-scientists do not always survive the trip.

George and Eva Klein

A few years ago, the British Psychoanalytical Society invited one of us to participate in a dialogue about biology. The first question was about apoptosis, or programmed cell death. "Would you not agree," went the question, "that the existence of a suicidal program that could be activated in any cell proves the ageing Freud's theory about a 'death wish' (*Todeswunsch*)?" I replied that apoptosis is more concerned with the life of an organism than with its death. A caterpillar would never turn into a butterfly, nor a tadpole into a frog, without extensive apoptosis in tissues that are no longer needed. We would all die of leukaemia or lymphoma at an early age if our lymphocytes — engaged as they are in a gigantic Monte Carlo game of immunoglobulin or T-cell-receptor rearrangement that allows us to react against a vast number of proteins — failed to eliminate the large majority that do not encounter their complementary antigen within a limited period of time. Apoptosis has nothing to do with the Freudian concept, except for sharing the misleading word 'death'.

Words tossed between C. P. Snow's 'two cultures' — scientists and non-scientists — may lead to more serious confusion than a little misunderstanding about apoptosis. Meeting a linguist friend at a social gathering, we came to speak about Luigi Cavalli-Sforza's studies on genetic markers in human populations and the derived conjectures about prehistoric migrations. Our friend said that he was well aware of this work and found it very beautiful, but that the conclusion that early human culture migrated with the genes was "unacceptable nonsense" because it could promote racism. I tried to explain that "culture" refers mainly to agriculture and that the absence of efficient information transfer across tribal and language barriers in prehistoric society made it understandable that new technologies arising in one location would be spread mainly by movements of people and therefore genes.

Our continued conversation made it clear, however, that my linguist colleague took Cavalli-Sforza's conclusion to imply that human culture is due to superior genes. He failed to see that the genetic markers used in these studies were merely flags. Many of them were not even genes, but polymorphic sites in non-coding DNA. My inability to convince my friend indicated to me that the Nazi mythology of superior and inferior

Words tossed between C. P. Snow's 'two cultures' may lead to serious confusion.

races may still cast its shadow over dialogue between biologists and non-biologists.

The notorious and much-abused misunderstanding between the two worlds revolves around the word 'chance'. Richard Dawkins's highly enjoyable and informative popular books about evolution, such as *The Blind Watchmaker*, have still left many of those who prefer to live in a pre-darwinian world unconvinced. They prefer the eighteenth-century theologian William Paley's dictum that no watch could be made 'by accident' and without a watchmaker. If they would only read Dawkins, they would see why the notion of darwinian evolution is 'counterintuitive'.

They might also consult a more recent book by the same author, *Climbing Mount Improbable*. In this book, Dawkins compares all existing organisms to mountain-tops. Each of them has attained its elevated position through a long series of mutations — incomprehensibly long for our subjective concept of time, which is moulded to fit our limited lifespan. Each new mutant had to pass through the needle's eye of selection, not once, but continuously over many millions of years. 'Mount Improbable' has a precipitous, almost perpendicular wall with forbidding cliffs to the north and gentle, grass-covered slopes to the south. Critics of evolution often argue against the role of 'chance', as if evolutionary theory would claim that each species reached the top by jumping from the plain to the peak in a single, randomly occurring leap. No biologist claims anything like that. They envisage a slow uphill walk along the gentle slope. The title of Jacques Monod's classical book *Chance and Necessity* catches the essence of the process. 'Chance' refers to random mutations, whereas 'necessity' represents the selective survival of the fittest. Mutations only provide the basis for what can happen; selection decides what actually prevails. To talk about chance alone is to miss the point.

Do biologists run the risk of serious misconceptions in their relationship with the humanities and the social sciences? In conversations with molecular biologists of



Peak of perfection: non-biologists do not always understand how organisms reach the summit.

different ages, we sometimes encounter the opinion that a detailed molecular understanding of the central nervous system and its function will render most of the social sciences and humanities, including linguistics, psychology and even poetry, superfluous. In our own conversations, we refer to this view as 'molecular fundamentalism'.

We know that all human culture stems from the activities of our neurons and the organization of our brain. But while engaging in the mandatory reductionistic analysis without which there can be no science, we must remember that the whole is more than the sum of its parts. Each new level of complexity has a distinctive world made up of its terminology, its way of reasoning, its *modus operandi*. We must resist the fundamentalistic temptation to believe that the 'words' of the system — that is, the molecules that provide the basis of its structure and function — can also explain the potentially infinite number of complex 'sentences' and other superstructures that our brain has built with them. ■

George and Eva Klein are at the Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm S-17177, Sweden.

HANS CHRISTOPH KAPPEL/BBC WILD

More than words

Geoffrey K. Pullum and
Barbara C. Scholz

In the popular view, a language is merely a fixed stock of words. Purists worry about foreign loanwords; conservatives decry slang; and groundless claims that there are hundreds of Eskimo words for snow are constantly made in popular writing, as if nothing matters about languages but their lexicons.

But the popular view cannot be right, because (as linguist Paul Postal has observed) membership in the word stock of a natural language is open. Consider this example: "GM's new Zabundra makes even the massive Ford Expedition look economical." If English had an antecedently given set of words, then this expression would not be an English sentence at all, because 'Zabundra' is not a word (we just invented it). Yet the sentence is not just grammatical English, it is readily interpretable (it clearly implies that the Zabundra is a large, fuel-hungry sports utility vehicle produced by General Motors). Similar points could be made regarding word borrowing, personal names, scientific nomenclature, onomatopoeisis, acronyms, loaned words, and so on; English is not a fixed set of words.

A more fundamental reason that a language cannot just be a word stock is that

expressions have syntactic structure. For example, in most languages, the order of words can be significant: "Mohammed will come to the mountain" contains the same words as "The mountain will come to Mohammed", but the expressions are very different. Inclusion within phrases is also an important part of syntactic structure. In the expression "We could not tell him", ambiguity arises from the fact that 'not' may belong in the same phrase as 'tell him' — in which case the meaning is that keeping him in the dark is possible or permitted — or it may be outside the 'tell him' phrase, in which case 'not' belongs with 'could' and the meaning is that telling him is impossible or forbidden.

The syntactic structure of natural languages has several important features. One is revealed by the example just considered; there is no guarantee of what mathematical logicians call 'unique readability' — there is no one-to-one correspondence between sound strings and syntactic structures, or between syntactic structures and meanings. Natural languages are replete with ambiguity.

A second feature is that there is no upper limit on the complexity of expressions. A verb phrase such as 'run away' can be embedded in a larger verb phrase such as 'see Spot run away', and there is no syntactic limit on further embedding, so expressions can be of arbitrary complexity: "Tell him they think he overheard someone ask her to confirm that they saw him watching us waiting for you to see Spot run away in order to ...". Hence, natural languages are productive, as they possess the structural resources for indefinite recombination.

A third feature of natural language syntax is its variability. Even within a single speech community (which we can roughly define as a human group whose members broadly understand each other's speech and recognize it as being characteristic of the group), there are quite sharp differences concerning the relevant regularities of syntactic structure, both between subgroups (dialect differences) and between individuals, whose idiosyncratic divergences mostly go unnoticed.

Fourth, malformations of syntax vary in their severity — some partially ill-structured expressions are more deviant than others. President George W. Bush's departures from standard English syntax are well known; Tarzan's departures are more extreme, and Yoda's even more so ("Already know you that which you need"); yet the structure of English is partially respected in each case. Likewise, familiar phenomena such as hesitation ("It's in the ... in the drawer") and use of fragments ("And that would be ...") partially conform with English expressions; they are not random jumbles of words.

Language

Natural languages are unique in combining the best of formal languages and primate communication.

Within mathematical logic and computer science, invented formal symbolic systems are called languages, but in some respects they are strikingly different. Their syntactic structure allows embedding and re-embedding, but their vocabularies are fixed; ambiguity is ruthlessly excluded; structures must be completely well-formed, and disrupted or fragmented expressions simply do not belong at all. Most of the linguistic work on syntactic theory over the past 50 years has used the mathematical methods devised for defining formal languages of this sort. This suggests a possible confusion of formal tools with subject matter.

To formulate grammars that describe natural languages precisely, we need a formal metalanguage that has the properties of any other scientific, formal language: a recursively defined, unambiguous syntax that defines a countably infinite set of expressions. But natural languages themselves — the focus of scientific linguistics — are not precisely delineated sets of expressions, any more than they are precisely delineated sets of words.

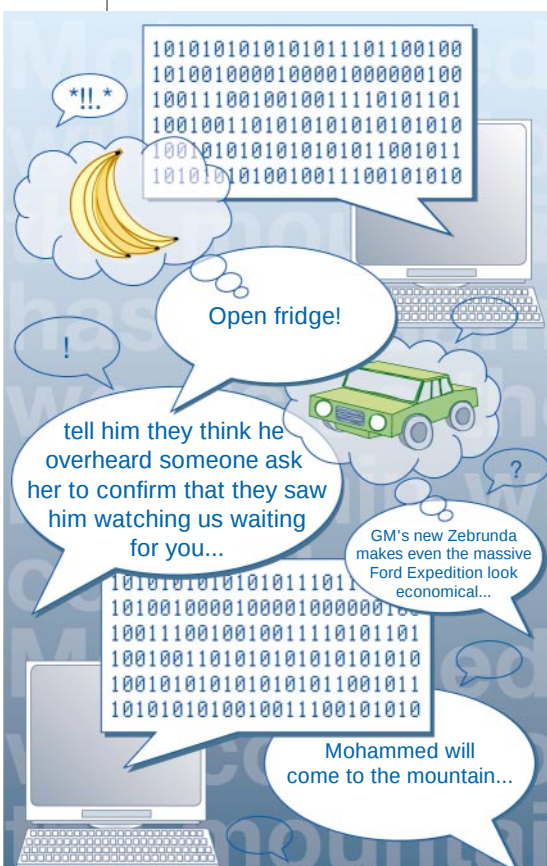
The aspects of language that can be learned by certain great apes, particularly the chimpanzees *Pan troglodytes* and *Pan paniscus*, are curiously the very same ones that fit the popular conception of languages. Apes can learn to name things using hand signs or visual symbols, and to express some basic demands by uttering them ("Open fridge! Give apple give!"), but they seem to be incapable of developing a productive grasp of syntax.

Human languages exhibit a unique combination of characteristics: first, semantic word-to-world relations that we share with other primates; second, syntactic structures as complex and exact as in formal languages; and third, an openness, flexibility and ambiguity that formal languages do not allow. ■

Geoffrey K. Pullum is in the Department of Linguistics, University of California, Santa Cruz, California 95064, USA. Barbara C. Scholz is in the Department of Philosophy, San José State University, San José, California 95192, USA.

FURTHER READING

Karmiloff, K. & Karmiloff-Smith, A. *Pathways to Language* (Harvard Univ. Press, Cambridge, MA, 2001). Pullum, G. K. & Scholtz, B. C. in *Logical Aspects of Computational Linguistics, 4th Int. Conf.* (eds de Groote, P., Morrill, G. & Retoré, C.) 17–43 (Springer, Berlin, 2001). Wallman, J. *Aping Language* (Cambridge Univ. Press, Cambridge, 1992).



Once upon an asteroid

Erik Asphaug

When the NEAR-Shoemaker spacecraft landed on the asteroid Eros earlier this year, it provided an unprecedented view of these battered relics from the early Solar System. The next step is to find out what's inside.

Planetary formation began as delicate sedimentation, with granules clinging to other granules in the solar nebula¹. It climaxed 10 million to 100 million years later with planets smashing together at tens of kilometres per second². The Moon, for instance, was created when a body the size of Mars hit the early Earth about 4.5 billion years ago^{3,4}. In between was a fast-paced epoch, as poorly comprehended as it was brief, which we try to understand by studying asteroids and comets — remnants that wander the Solar System like ghosts from a bygone time. One such ghost is 433 Eros, and papers beginning on page 390 of this issue^{5–7} describe the scientific results from the controlled descent of NASA's NEAR-Shoemaker spacecraft onto this asteroid last February.

Named after planetary geologist Eugene Shoemaker, the Near Earth Asteroid Rendezvous (NEAR) spacecraft was the first to fly under NASA's 'faster, better, cheaper' Discovery programme. After a flyby of main-belt asteroid 253 Mathilde⁸ in 1997, NEAR-Shoemaker became the first spacecraft to orbit an asteroid. Although it was not designed to land on Eros, after one year it began its descent at walking pace (Fig. 1) and relayed to Earth a tantalizing sequence of final images: block-strewn landscapes, pitted rocks and mobile soils. Camera transmissions ended when the spacecraft hit the surface, but the γ -ray spectrometer continued for days to collect data on the elemental composition of Eros. Thus ended our first detailed investigation of a near-Earth object — a category that includes the potentially hazardous asteroids whose orbits may one day coincide with that of the Earth. Interest in these bodies goes well beyond the purely scientific, and includes investigation of ways in which asteroids could be deflected from a collision course with Earth, or mined for their unique and abundant resources^{9,10}.

Eros originated in the main asteroid belt, beyond the orbit of Mars, where Jupiter's gravitational stirring stunted planet growth. The belt is crowded, and the long-term fate of asteroids is to batter each other to bits. Relics do survive; Mathilde is one. A variety of dynamical processes cause objects to leak from the main belt into the inner Solar System where they typically crash into the Sun or a planet, or are ejected from the Solar System, after about ten million years¹¹. Planet-crossing orbits tend to be chaotic, so fates

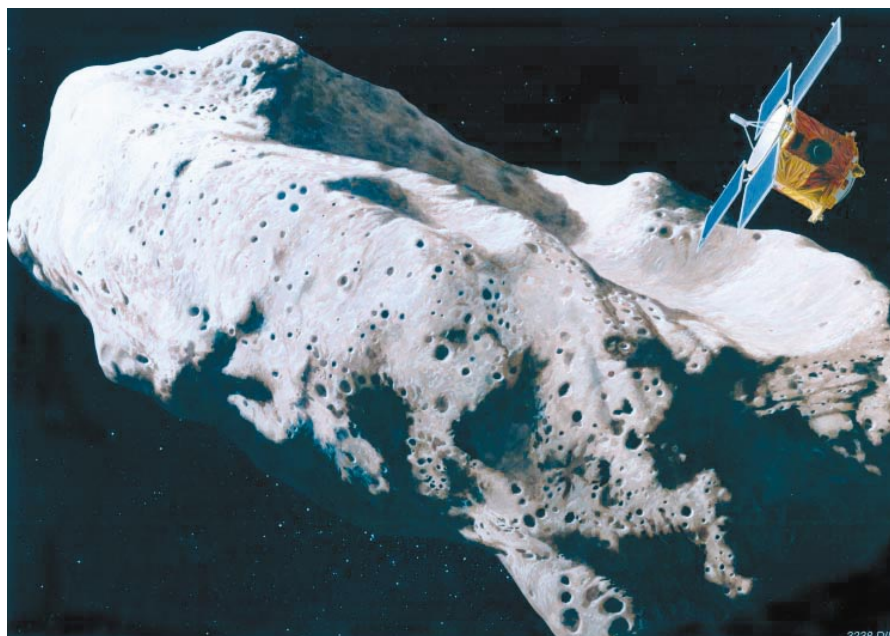


Figure 1 Going down: an artist's impression of NEAR-Shoemaker approaching Eros.

are expressed as probabilities: Eros has a chance of around 5% of striking Earth¹², but not anytime soon.

Eros is about $34 \times 13 \times 13$ km in size, and is the second largest of the known near-Earth objects; the largest is 1036 Ganymed, which is some 40 km in diameter. Eros bears enormous impact scars from its four billion years in the main belt, and may have lost much of its original mass (and shape) to cratering. But X-ray and γ -ray measurements have indicated¹³ that it has a relatively unaltered, primitive composition: it was probably never part of a much larger parent body that differentiated, by heat and other factors, into compositionally different parts such as a mantle and a core.

In the papers in this issue, Veverka *et al.*⁵ discuss the images obtained during the spacecraft's descent. Robinson *et al.*⁷ describe an asteroid surface that is awash in sediment, featuring strikingly flat 'ponds' within many craters. And Thomas *et al.*⁶ show how thousands of large blocks of material probably rained down as impact ejecta from a relatively young eight-kilometre crater, provisionally named Shoemaker. It was once believed that impact pounding of small asteroids would scour the surface clean of fragmented material (regolith) because any material moving faster than a few metres

per second will escape the asteroid's tenuous gravity (only 0.06% of that of Earth) and disappear into space¹⁴. But data from NEAR-Shoemaker reveal that sedimentation and gravitational re-accumulation processes have dominated the surface layers.

If the largest boulders came from a recent crater, as Thomas *et al.*⁶ propose, then Eros is an impact-modeller's delight. A successful model would require all of the boulders to be ejected at the right speed while making the right size and shape of crater. Such simulations might tell us about the interior properties of Eros. If the interior were modelled as too resistant to fragmentation, for example, then ejecta blocks would be blasted out too fast because of the intense excavational energy; if too easily fragmented, the blocks would be too small. A remaining puzzle is the near-absence of pits around the boulders, which should have formed in these loose and dusty soils upon the blocks' re-entry at several metres per second.

The ponded sediments are a mystery whose solution may require a better understanding of particulate dynamics under microgravity conditions. Robinson *et al.*⁷ propose that photoionization of sub-micrometre-sized grains in the sunlit equatorial latitudes of Eros has resulted in the grains' charged, electrostatic levitation¹⁵, winnowing

NASA/JPL

them out of the upper regolith. According to this explanation, the particles then settled, fluid-like, into gravitational lows such as impact craters. The ponds' slightly blueish colour may be attributed to the optical properties of submicrometre particles. Alternatively, particle sorting might be controlled by composition, with different minerals migrating towards the ponds¹⁶. Seismic shaking is another proposed mechanism for soil transport on asteroids. All in all, asteroids could be the perfect low-gravity laboratory for advancing the physics of granular media. Such media — agricultural grains, pharmaceutical powders, ores and chemicals — are the basis for hundred-billion-dollar industries, but their behaviour continues to baffle physicists and engineers¹⁷.

If charged dust is a large component of asteroid regolith, it will be a real problem for both human and robot explorers. Each minor disturbance would send dust clouds aloft, like mud from the bottom of a lake, and the particles would cling to and penetrate equipment, and could eventually enshroud the asteroid in a transitory 'atmosphere'. Lunar dust was a serious difficulty for the Apollo astronauts, so the behaviour of charged dust in microgravity must be studied before more ambitious exploratory activity is attempted on an asteroid, and certainly before the first crewed landing. For example, the Japanese mission MUSES-C will hover above the surface of the near-Earth asteroid 1998 SF36, which is 500 metres in diameter, and use a bullet to knock off a sample to bring home in 2007 — essentially fulfilling Shoemaker's dream to bang on an asteroid with a rock hammer. More complex missions, including one to perform small-scale blast experiments plus radar tomography and seismology, are on the drawing board¹⁸.

Suppose you could dig beneath the soil and boulders of Eros without creating too much mess. What would you find? Many suspect that the asteroid is soil and boulders all the way through, a hypothesis that might explain how it retains so much loose material on its surface: weak aggregates can absorb and dissipate shock energy¹⁹, which increases their chances of staying intact and so of accumulating regolith. On the other hand, regional and local faults and fractures suggest that the interior of Eros has been fragmented by impacts²⁰. The two views are not incompatible. The emerging picture is that Eros is indeed shattered through-and-through, but that these fragments formed in place without being shaken up into more loosely packed configurations with lower overall density.

Despite NEAR-Shoemaker's remarkable success, we remain profoundly ignorant about many aspects of asteroids. They are more subtle than we had imagined, and there is no telling how they would respond to the

more forceful poking and prodding that would be needed to divert them from a collision course with Earth. The cost and political complexity of defending against near-Earth objects requires a guarantee that a chosen technique will work, especially one involving space-borne thermonuclear devices — a proposed solution which may present its own dire hazard. At the moment the world remains preoccupied by other threats, but we should not forget that our future well-being depends upon becoming familiar with the inner workings of near-Earth asteroids. ■

Erik Asphaug is in the Department of Earth Sciences, University of California, Santa Cruz, California 95064, USA.

e-mail: asphaug@es.ucsc.edu

1. Weidenschilling, S. J. & Cuzzi, J. N. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1031–1060 (Univ. Arizona Press, Tucson, 1993).
2. Wetherill, G. W. *Science* **228**, 877–879 (1985).
3. Cameron, A. G. W. & Ward, W. R. *Lunar Planet. Sci. Conf.* **7**, 1204–1207 (1976).
4. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).

5. Veeverka, J. *et al.* *Nature* **413**, 390–393 (2001).
6. Thomas, P. C., Veeverka, J., Robinson, M. S. & Murchie, S. *Nature* **413**, 394–396 (2001).
7. Robinson, M. S., Thomas, P. C., Veeverka, J., Murchie, S. & Carcich, B. *Nature* **413**, 396–400 (2001).
8. Veeverka, J. *et al.* *Science* **278**, 2109–2111 (1997).
9. Lewis, J. S. *Mining the Sky: Untold Riches from the Asteroids, Comets and Planets* (Addison-Wesley, Reading, MA, 1996).
10. <http://www.nearearthobjects.co.uk> Report of the UK Task Force on Near Earth Objects (British National Space Centre, 2000).
11. Gladman, B., Michel, P. & Froeschlé, C. *Icarus* **146**, 176–189 (2000).
12. Michel, P., Farinella, P. & Froeschlé, C. *Astron. J.* **116**, 2023–2031 (1998).
13. Trombka, J. I. *et al.* *Science* **289**, 2102–2105 (2000).
14. Veeverka, J., Thomas, P., Johnson, T. V., Matson, D. & Housen, K. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 342–402 (Univ. Arizona Press, Tucson, 1986).
15. Lee, P. C. *Icarus* **124**, 181–194 (1996).
16. Akridge, D. G. & Sears, D. W. G. *J. Geophys. Res.* **104**, 11853–11864 (1999).
17. Jaeger, H. M., Nagel, S. R. & Behringer, R. P. *Physics Today* **49**, 32–39 (1996).
18. Asphaug, E., Belton, M. J. S. & Kakuda, R. Y. *Lunar Planet. Sci. Conf.* **32**, Abstr. 1867 (2001).
19. Love, S. G., Hörz, F. & Brownlee, D. E. *Icarus* **105**, 216–224 (1993).
20. Prockter, L. M. *et al.* *Lunar Planet. Sci. Conf.* **32**, Abstr. 1947 (2001).

Human genetics

Testing telomerase

Robert Marciniak and Leonard Guarente

The molecular defect in the human disease dyskeratosis congenita is a reduction in function of telomerase. This discovery provides a direct test of the importance of this enzyme in ageing and cancer.

Dyskeratosis congenita is an inherited human disease from which sufferers die between the ages of 16 and 50. Problems tend to occur in tissues in which cells multiply rapidly — skin, nails, hair, gut and bone marrow — with death usually occurring as a result of bone-marrow failure. The disease is inherited in one of two ways. Some people carry mutations in a gene, identified a few years ago¹, on the X chromosome. In others, mutations occur in unknown genes on non-sex chromosomes (autosomes).

The gene that is mutated in the X-linked disorder encodes the protein dyskerin¹, which is found in a subsection of the cell's nucleus known as the nucleolus. Although at first it was proposed that nucleolar dysfunction underlies the disease, it was later discovered² that mutations in dyskerin actually affect the RNA part of telomerase — an enzyme with both RNA and protein portions that helps to maintain the ends of chromosomes (telomeres). This suggested that dyskeratosis congenita might instead result from a defect in telomerase activity. On page 432 of this issue, Vulliamy and colleagues³ confirm this with their discovery that an autosomal dominant form of the disease is caused by mutations in the gene that encodes the RNA part of telomerase itself.

This question of which molecular defect

underlies dyskeratosis congenita has been difficult to answer. In people with the X-linked form of the disease, levels of telomerase RNA are low, and telomeres are shorter than normal in white blood cells and fibroblast cells². These findings support a role for dyskerin in stabilizing telomerase RNA and telomerase activity, and therefore support the idea that dyskeratosis congenita results from defective telomerase. But, as dyskerin is found in the nucleolus, a contribution from nucleolar dysfunction could not be ruled out.

The results of Vulliamy *et al.*³ seem to settle the matter. Using a combination of genetic approaches, the authors show that affected people in one family lack 821 bases from one of the two copies of the gene that encodes telomerase RNA (they are heterozygous for this deletion). The deletion removes the 74 bases at the end of the gene's coding region. In two other, unrelated families, affected people are heterozygous for either one or two single-base mutations in the telomerase-RNA gene. None of these mutations was found in the corresponding genes from 50 unaffected people.

So dyskeratosis congenita provides a test of the significance of telomerase defects in humans. This is interesting because it is thought that telomere loss might contribute to normal ageing. Every time a cell divides,

Organ system	Cells expressing telomerase	Defect in dyskeratosis congenita
Hair	Hair follicle	Alopecia
Oral cavity	Squamous epithelium	Leukoplakia (precancerous oral lesions)
Skin	Basal layer of epidermis	Abnormal pigmentation Nail dystrophy
Lungs	Type 2 alveolar epithelial cells	Fibrosis
Liver	?	Cirrhosis
Intestine	Intestinal crypts	Gut disorders
Testes	Spermatogonia	Hypogonadism
Bone marrow	Progenitor stem cells	Failure to produce blood cells

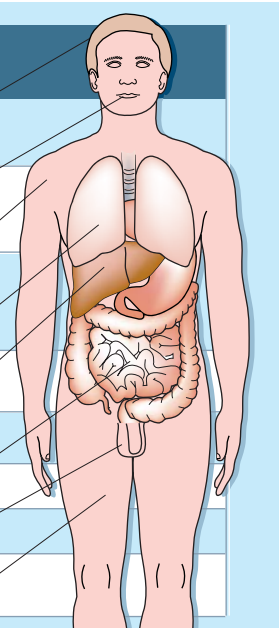


Figure 1 Characteristics of the human disease dyskeratosis congenita. Defects are seen most often in tissues in which cells divide rapidly and often. Many of these cells express telomerase, an enzyme that maintains telomeres (the ends of chromosomes), and Vulliamy *et al.*³ have shown that one form of dyskeratosis congenita can be caused by mutations in the gene that encodes the RNA component of telomerase.

it needs to replicate its DNA. But, because of the directional nature of DNA synthesis, the telomeres that cap the ends of linear chromosomes are incompletely replicated, and are eroded a little with each cell division. Certain mechanisms can compensate for this loss — for example, telomerase adds sequence back to chromosome ends.

Not all cells express telomerase, however. For instance, human fibroblasts do not, and have a finite ability to replicate in culture. This phenomenon, termed replicative senescence, is caused by progressive telomere shortening⁴. The importance of this process to the ageing of whole organisms has been unclear, but it was first tested in mice by disrupting the gene that encodes the RNA part of telomerase, creating animals with undetectable telomerase activity⁵. Early generations of these mice developed without significant abnormalities; defects were seen only in later generations, mostly in highly proliferative tissues^{6,7}, in which the erosion of telomeres is expected to be more severe.

If dyskeratosis congenita indeed results from telomerase deficiency, one would expect that people with this disease and the late-generation telomerase-deficient mice would have similar defects. This has proved correct — both show abnormalities in the production of blood cells and in the gut, as well as poor wound healing. This is despite the fact that mice lacking telomerase RNA have no detectable telomerase activity, whereas people with the autosomal form of dyskeratosis congenita do — there is at most a partial loss of function. Moreover, the fea-

tures of the X-linked form of dyskeratosis congenita are seen even in the first generation, although this could be because human telomeres are much shorter than those of laboratory mouse strains.

But do the symptoms of the mutant mice and of people with dyskeratosis congenita shed any light on the involvement of telomere shortening in human ageing? These mice and the human patients have only a few features — such as premature greying of hair and higher incidence of cancer — that are also seen as humans age. If telomere loss during ageing is a key determinant of the age-related decline in tissue function, why aren't more stereotypical features of premature ageing seen? The answer may lie in the tissue-specific regulation of telomerase expression (Fig. 1).

Telomerase activity has been detected in precursor (progenitor) cells in tissues that have high rates of cell turnover, including those of the blood system and the intestinal crypts. By contrast, some tissues that have the capacity for cellular replacement, but do not undergo continuous cell turnover, do not express telomerase in their progenitors. It is these tissues — such as the deep layers of the skin or the lining of the blood vessels — that might be expected to suffer most from age-associated telomere depletion, as they have no ability to regenerate telomeres. These tissues would also be greatly affected by defects in other pathways that maintain telomeres, such as DNA-recombination processes. This might explain why Werner's syndrome, in which an enzyme involved in DNA processing is affected, yields a closer



100 YEARS AGO

Among the local fêtes of parts of the north of France the procession of giants forms the most original and picturesque custom. Each Flemish town formerly possessed its giant, but this curious custom preserves its ancient ceremonial in only a few localities. Lille has not seen for a long time the procession of the giant "Phinar", which was vilified as was its colleague "Annéen" at Valenciennes. The festivals of giants are still preserved at Dunkirk, where "Papa-Rœusse" is the idol of the inhabitants, at Cassel, at Gand, at Brussels, and especially at Douai, where every June "Gayant" has a triumphal procession accompanied by his wife, "Marie Cageon", and their three children, "Jacquot", "Fillon" and "Bimbin". From *Nature* 26 September 1901.

50 YEARS AGO

In describing and assessing the results of artificial rain-making experiments, which attempt to induce precipitation by 'seeding' suitable clouds with pellets of dry ice, crystals of silver iodide or small water-droplets, Mr. Mason said that there has been a distinct tendency to draw spectacular conclusions on the basis of too few observations, and that too little attention has been paid to the need for adequate control experiments. However, many cases have been reported of rain falling from clouds a few minutes after seeding. Although in any particular case it is not possible to ascertain that this would not have happened without seeding, in many careful dry-ice experiments carried out in Australia... rain has been observed to fall from seeded cumulus clouds, while similar clouds in the immediate vicinity produced no precipitation. Since, for a high probability of success in the dry-ice experiments, the cloud top must be colder than -7°C . and rain is likely to fall naturally if it is colder than -12°C ., there is only a narrow range of conditions in which seeding with dry ice can anticipate natural events. The position is still more unfavourable in the case of silver iodide, which suffers the additional disadvantage that it becomes inactive as an ice nucleus when exposed to strong sunlight for about one hour. Calculations indicate that spraying small water-droplets into the base of the cloud should be a more efficient method of releasing showers from warm cumuli, and the results of recent experiments in Australia are encouraging. From *Nature* 29 September 1951.

version of normal (if premature) ageing than does dyskeratosis congenita. In people with dyskeratosis congenita and in telomerase-deficient mice, it is tissues that normally express telomerase that one would predict to suffer most from its loss, and this proves to be the case.

People with dyskeratosis congenita, as well as late-generation telomerase-deficient mice, also suffer from a higher rate of cancer. This can likewise be explained by the lack of telomerase, which results in unstable chromosomes — in dyskeratosis congenita sufferers and the mutant mice, many chromosomes fuse end-to-end^{8,9}, probably because their telomeres are terminally eroded (discussed in ref. 10).

In short, the symptoms of this disease provide a glimpse of the effects of telomerase defects on the maintenance of human tissues. Cells in rapidly dividing tissues, with progenitors that usually express telomerase, are more strongly affected. It will be inter-

esting to see whether Werner's syndrome and related premature-ageing disorders affect the maintenance of telomeres in tissues that do not have telomerase-expressing progenitors.

Robert Marciniak is in the Departments of Medicine and of Cellular and Structural Biology, University of Texas Health Science Center, San Antonio, Texas 78229, USA.

Leonard Guarente is in the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mails: marciniak@oncology.uthscsa.edu
leng@mit.edu

1. Heiss, N. S. *et al. Nature Genet.* **19**, 32–38 (1998).
2. Mitchell, J. R., Wood, E. & Collins, K. *Nature* **492**, 551–555 (1999).
3. Vulliamy, T. *et al. Nature* **413**, 432–435 (2001).
4. Bodnar, A. *et al. Science* **279**, 349–352 (1998).
5. Blasco, M. A. *et al. Cell* **91**, 25–34 (1997).
6. Rudolph, K. L. *et al. Cell* **96**, 701–712 (1999).
7. Herrera, E. *et al. EMBO J.* **18**, 2950–2960 (1999).
8. Artandi, S. E. *et al. Nature* **406**, 641–645 (2000).
9. Dokal, I. *Br. J. Haematol.* **110**, 768–779 (2000).
10. de Lange, T. & Jacks, T. *Cell* **98**, 273–275 (1999).

Statistical physics

The salesman and the tourist

H. Eugene Stanley and Sergey V. Buldyrev

Solutions to optimization problems, such as that faced by the travelling salesman, have many practical applications. Might a related problem offer insight into the behaviour of foraging animals?

The travelling salesman problem is of fundamental interest to mathematicians and physicists, and has a number of practical applications, such as computer design. At first sight, the related travelling tourist problem might seem to be of little interest to anyone except perhaps quarrelling travelling companions.

What are these two problems? The travelling salesman problem is the question of what is the shortest possible route followed by a salesman assigned to visit N clients spread across a huge territory; it is a classic example of a class of problems called global optimization problems. Because the salesman's next step depends on his knowledge of the location of every client yet to be visited, the solution to the travelling salesman problem is far from trivial, and has attracted some of the best minds^{1–3}.

In *Physical Review Letters*, Lima *et al.*⁴ introduce a new example from a class of problems known as local optimization problems, and name it the travelling tourist problem. This is the question of what is the best path to be followed by a tourist not constrained to only one visit per city, but constrained — say by his budget — to visit the nearest city. Note that the tourist wants to minimize only the distance to the next city, a local optimization, not the sum of all the distances along the trajectory as in the

travelling salesman problem. So, at first sight, the travelling tourist problem would seem to be trivial.

Nonetheless, even a trivial-sounding problem can yield rich behaviour. For example, the random-walk problem considers a walker limited to the vertices of a square grid. At each step the walker moves to the neighbouring north, south, east or west vertex depending on the throw of a four-sided die. Clearly the walker does not need to know more than his immediate environment. Even so, his behaviour cannot be predicted in many cases, such as when he

cannot revisit a vertex, or if the grid has some randomly scattered sites that can trap the walker.

Similarly, the travelling tourist problem exhibits some surprising behaviour. For example, after visiting a few cities the tourist will visit city A, whose nearest neighbour, city B, has no other neighbour closer than A. So as soon as the tourist visits either A or B, he will thereafter simply oscillate between them, in what is called a two-city limit cycle. Such two-cycles act as effective 'tourist traps' with the tourist going back and forth, like Sisyphus, for eternity.

Lima *et al.*⁴ enrich the travelling tourist problem by assuming that the idealized tourist is more like a real tourist and does not wish to visit the city he went to last, so they introduce a rule that prevents this. The same sort of logic applies as before, except that now the tourist can be trapped in a three-cycle (three cities that form a triangle A-B-C such that a tourist starting at A will pass to B and then to C, before returning to A). Surprisingly, in addition to three-cycle traps, there are also p -cycle traps, where p is as large as desired.

This simple rule can be further generalized if the tourist cannot visit a city already visited in the previous V visits (in the two previous cases, $V=0$ and 1, respectively). Remarkably, after a trip exploring several cities, the tourist is always trapped into repeatedly revisiting a subset of them. Indeed, the entire set of N cities can be partitioned into a large number of 'free cities' that the tourist can enter and then leave, and a number of tourist traps, in which he is condemned to revisit repeatedly the same subset of p cities (Fig. 1).

So, at first sight, the travelling tourist confronts a landscape not so different from that confronted in the random-walk problem with traps. But in the random-walk problem, the traps are typically assumed to have sizes that follow a normal distribution centred around a typical size. In the travelling tourist problem, the p -cycle traps can have a range of p -values, so the number of

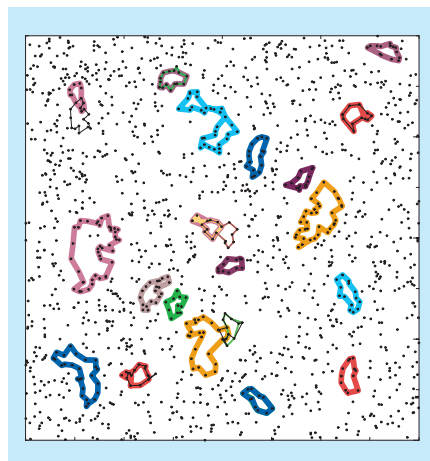


Figure 1 The travelling tourist problem. An example in which the tourist is allowed to visit any one of 1,600 random 'cities', subject only to the condition that he does not revisit a city visited in the previous V visits, where $V=10$. As well as a great number of free-standing cities to which the tourist can come and go, there are 378 cities that belong to one or more p -cycles, which serve as 'tourist traps' by constraining the tourist to visit the same set of cities over and over again. The 29 traps shown here occur in a range of sizes and shapes. Lima *et al.*⁴ show that the distribution of these p -cycles follows a power-law function. Such a distribution is normally associated with random processes, rather than the completely deterministic walk followed by the tourist.



Figure 2 A wandering albatross looking for food. This bird may follow a stable migratory route — it visits the same sites (typically feeding areas) over and over again in much the same way as the hapless tourist in Fig. 1. (Reproduced with permission of the British Antarctic Survey.)

traps of a given size might be expected to follow a normal distribution. But Lima *et al.*⁴ find that the number of p -cycles follows a monotonically decreasing power-law function of p , with an exponent of about 2.7. This is a surprising result because power laws are typically associated with systems dominated by random processes, but here the walk is completely deterministic. The power law must somehow arise from the initial random distribution of the cities, so it is intriguing to think there may be deep similarities between this problem and typical power-law systems⁵.

There are many examples of trivial-sounding problems having unexpected applications; for example, random-walk problems can describe a rich range of diffusive motion⁶, and are even used as a crude model of price movements on the stock market⁷. Accordingly, the travelling tourist problem could apply to some interesting questions. Some animals such as birds follow migratory routes in which they revisit an identical set of feeding spots in the same sequence (Fig. 2). These animals may not be altogether unlike the tourist, choosing deterministically the closest feeding ground to where they are, and conditioned by the fact that if they revisit the same feeding area without waiting a certain time the food supply will not have had the chance to rejuvenate itself. So the travelling tourist might provide new insights into the foraging behaviour of insects, mammals and birds, which can follow power-law functions^{8,9}.

Many humans tend, after some time, to fall into a rut — to revisit the same sequence of cities on their travels, to revisit the same sequence of restaurants in their social life, and even for some scientists to revisit the same set of ideas in their scholarship — all actions seemingly occurring as if by magic when in fact there may be a simple underlying reason. Indeed, this type of behaviour is more like the travelling tourist problem than the travelling salesman problem. Normally we think of our behaviour as emerging

from free will. It is possible that some behaviour patterns are the result of applying deterministic rules to a stochastic (random) envi-

Plant hormones

Transporters on the move

Mark Estelle

The hormone auxin is moved across plant cells by transporters called efflux carriers. It now looks as if these carriers behave much more dynamically than had been thought.

If you're a plant you cannot get by without auxin, for it regulates various essential aspects of growth and development. In their paper on page 425 of this issue¹, Geldner *et al.* provide a surprise for biologists studying auxin transport. But their results will also be of wider interest, for they bear more generally on the questions of how cell polarity is established and maintained, and on protein trafficking in plant cells.

Auxin is a hormone that operates through its effects on cell division and elongation. It is transported through files of cells by a process that is thought to depend on the asymmetric distribution of auxin 'efflux carriers' acting at a cell's plasma membrane. The importance of this transport system has been amply demonstrated using synthetic inhibitors of polar auxin transport, which were believed to interfere specifically with the action of the efflux carrier. Geldner *et al.* now show that the efflux carriers cycle rapidly between the plasma membrane and an as-yet-unspecified cellular compartment, and that the inhibitors prevent this cycling. Further, they find that auxin-transport inhibitors are not specific: they have a general effect on the transport of membrane proteins.

Auxin is produced in young organs, primarily near the plant apex. It is then transported to other parts of the plant by influx and efflux carriers, located at the plas-

ma membrane, perhaps not altogether unlike the rules governing the journey of the travelling tourist.

H. Eugene Stanley and Sergey V. Buldyrev are at the Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA.

e-mails: hes@bu.edu

buldyrev@bu.edu

1. Kirkpatrick, S., Gelatt, G. D. & Vecchi, M. P. *Science* **220**, 671–680 (1983).
2. Baskaran, G., Fu, Y. & Anderson, P. W. *J. Stat. Phys.* **45**, 1–25 (1986).
3. Mézard, M. & Parisi, G. *Europhys. Lett.* **2**, 913–918 (1986).
4. Lima, G. F., Martinez, A. S. & Kinouchi, O. *Phys. Rev. Lett.* **87**, 010603 (2001).
5. Stanley, H. E. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
6. Ben Avraham, D. & Havlin, S. *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge Univ. Press, 2000).
7. Mantegna, R. N. & Stanley, H. E. *Introduction to Econophysics: Correlations & Complexity in Finance* (Cambridge Univ. Press, 2000).
8. Viswanathan, G. M. *et al. Nature* **401**, 911–914 (1999).
9. Viswanathan, G. M. *et al. Nature* **381**, 413–415 (1996).

ma membrane, which move the hormone through files of cells by successive rounds of transport into and out of these cells². The key to this mechanism is the asymmetric distribution of the efflux carrier. For example, auxin in the stem is transported in a highly directional fashion from the top towards the base of the plant. The PIN1 protein is a member of a large family of putative efflux carriers, and in the model plant *Arabidopsis* (and by implication all plants) it is required for this polar transport of auxins. Various studies in different plant organs found, in each case, that the efflux carrier is located on the downstream side of the transporting cell^{3–5}.

Our understanding of auxin transport comes largely from studies using artificial inhibitor compounds^{2,6}. Two of the best-characterized are called TIBA and NPA. These compounds reduce polar auxin transport in stem segments by blocking cellular auxin efflux, implying that they specifically interfere with the efflux carriers. The biological effects of transport inhibition are diverse and dramatic. Treatment of developing embryos with NPA causes defects in pattern formation and behaviour of the meristems, the groups of stem cells that give rise to all postembryonic structures⁷. Later in development, the inhibitors prevent proper growth of roots and shoots in response to environmental cues such as light or gravity⁸. Bio-

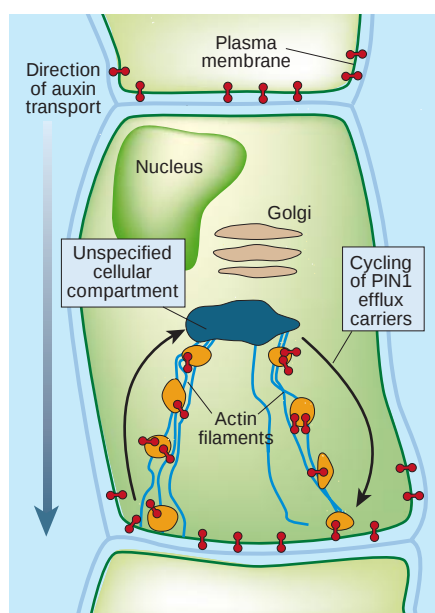


Figure 1 Auxin transport and PIN1 protein in a plant cell. PIN1 (red) acts as an auxin efflux carrier, and the generally accepted view has been that the carrier is sited statically at the cell's plasma membrane. But Geldner *et al.*¹ show that PIN1 seems to cycle rapidly between the plasma membrane and some unspecified compartment in the cell. Actin filaments are implicated in the cycling process; so far, the Golgi apparatus is not. Geldner *et al.* also show that certain synthetic compounds that have been taken to be specific inhibitors of auxin transport, including TIBA and NBA, actually affect protein transport more generally. This will have to be taken into account in auxin-inhibition studies.

chemical studies suggest that the inhibitors bind an unknown protein (the NPA-binding protein or NBP) that is associated with the membrane fraction but is distinct from the efflux carrier^{2,6}. Hence the view that the transport inhibitors are likely to interact with a regulator of auxin efflux.

Geldner *et al.*¹ were following up the observation⁹ that the *gnom* mutants of *Arabidopsis* are deficient in the asymmetric localization of PIN1. *GNOM* encodes a regulator of vesicle transport that is sensitive to the drug brefeldin A. Not surprisingly, brefeldin A also prevents normal localization of PIN1. In roots that have been treated with the drug, PIN1 accumulates in an undefined intracellular compartment. More surprisingly, this effect was rapidly reversible, even when protein synthesis was completely inhibited.

The implications are that PIN1 is relatively stable but rapidly cycles between the plasma membrane and an internal compartment (Fig. 1). The mechanism is unclear, but Geldner and colleagues provide some clues to it. First, they show that PIN1 accumulates in the intracellular compartment before there is much effect on the Golgi stacks. The implica-

tion is that PIN1-containing vesicles are not derived directly from the Golgi, which, as a central part of a plant cell's secretory apparatus, might have been expected to be involved. Second, Geldner *et al.* find that PIN1 cycling requires actin filaments, which is perhaps predictable given actin's involvement in vesicle movement. Treating roots with compounds that depolymerize actin prevented brefeldin-A-induced accumulation of PIN1 in the intracellular compartment and also prevented movement back to the plasma membrane when the drug was removed. Thus PIN1 cycling is actin dependent.

So how do the auxin-transport inhibitors affect this process? TIBA had little effect on PIN1 localization by itself, but it did block brefeldin-A-induced changes in that process. Further, adding TIBA after brefeldin-A treatment prevented PIN1 relocalization to the plasma membrane when brefeldin A was washed out. Thus TIBA inhibits the movement of PIN1 from, and back to, the plasma membrane. To see if this effect is specific, Geldner *et al.* tested the effects of brefeldin A and TIBA on two other plasma membrane proteins: an H^+ -ATPase and a syntaxin. Brefeldin A affected localization of the proteins, but in each case TIBA prevented these effects. Similar results were obtained with three other transport inhibitors, although their effective concentrations differed.

What should we conclude from these results? First, it is quite clear that TIBA and other so-called specific inhibitors of auxin efflux have a much more general effect on

protein transport. How they work and what this may ultimately tell us about protein trafficking remains to be seen. Actin filaments may be particularly important, as earlier studies showed that NPA binding to the NBP was dependent on them¹⁰. Perhaps the inhibitors disrupt an association between certain vesicles and the filaments. In any case, we need to keep these new results in mind when using the inhibitors to study auxin biology.

Second, localization of PIN1 to the plasma membrane appears to be a highly dynamic process. Given that TIBA inhibits both PIN1 cycling and auxin efflux, it seems that cycling may well be an essential aspect of PIN1 function. If this possibility is confirmed, we will have to revise our view of the efflux carrier as a static auxin transporter and rethink auxin transport and its regulation.

Mark Estelle is in the Section of Molecular Cell and Developmental Biology, Institute of Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: mestelle@mail.utexas.edu

1. Geldner, N., Friml, J., Stierhof, Y.-D., Jürgens, G. & Palme, K. *Nature* **413**, 425–428 (2001).
2. Palme, K. & Galweiler, L. *Curr. Opin. Plant Biol.* **2**, 375–381 (1999).
3. Galweiler, L. *et al.* *Science* **282**, 2226–2230 (1998).
4. Gil, P. *et al.* *Genes Dev.* **15**, 1985–1997 (2001).
5. Muller, A. *et al.* *EMBO J.* **17**, 6903–6911 (1998).
6. Estelle, M. *Plant Cell* **10**, 1775–1778 (1998).
7. Hadfi, K., Speth, V. & Neuhaus, G. *Development* **125**, 879–887 (1998).
8. Estelle, M. *Curr. Biol.* **6**, 1589–1591 (1996).
9. Steinmann, T. *et al.* *Science* **286**, 316–318 (1999).
10. Butler, J. H., Hu, S., Brady, S. R., Dixon, M. W. & Mudge, G. K. *Plant J.* **13**, 291–301 (1998).

Quantum physics

Entangled atomic samples

J. Ignacio Cirac

Quantum mechanics has potential applications in communication and computation. But first a quantum connection — known as entanglement — has to be created between bigger and bigger objects.

Many of the technological developments of the past century are based on the principles of a strange theory called quantum mechanics. This is true for lasers, nuclear reactors, semiconductors and many other systems that have changed our lives. Apart from these technologies, the advent of quantum mechanics introduced a new way of looking at nature, attracting the attention of philosophers as well as scientists. Quantum mechanics predicts the existence of entangled states that have counterintuitive properties. These states can involve two or more objects, and their unusual properties mean that when we perform a measurement on one object, the others become instantaneously modified, wherever they are. Entangled states are needed for teleportation experiments, in which one can trans-

fer the (quantum) properties of one object to another in a different location, and in quantum communication and computation.

To date, entangled states of a few atoms and photons have been prepared in a number of experiments. But some applications need much bigger entangled objects, formed by a large number of atoms. On page 400 of this issue, Julsgaard *et al.*¹ report an experiment in which two separate samples of atoms, each sample containing trillions of atoms, are entangled. The methods they use are attractive because they may lead to improvements in certain aspects of quantum communication and computation. This work should also pave the way for a new generation of experiments that allow researchers to teleport states of matter^{2–5}.

According to quantum mechanics, the

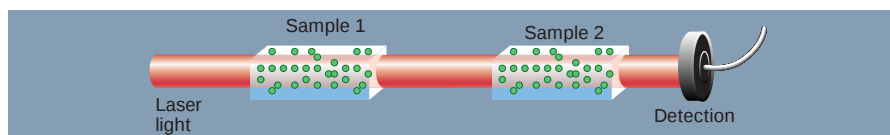


Figure 1 Method used by Julsgaard and colleagues¹ to entangle two atomic samples. The same laser light is sent through both atomic samples. Its frequency and polarization are carefully selected so that after the interaction the spin of the atoms in both samples becomes entangled with the laser beam. A suitable measurement of the polarization of the light then changes the atomic state in such a way that it becomes a superposition of states that are compatible with the outcome of the measurement. These states are symmetric with respect to the quantum properties of the atoms in each sample — practically half the atoms are spin up and the other half spin down — which protects the entangled state against destructive interactions with the environment.

properties of an object are not necessarily well defined if we do not look at them. Suppose I have a coin in my hand and then, when I look at it, I see that it lies tails up. This does not necessarily mean that the property of being heads up or tails up was defined before my observation. As long as we do not 'observe', there exist intermediate situations between being tails and heads up (something like being part heads up and part tails up at the same time), which are called quantum superpositions. Of course, we cannot directly see these superpositions because, as soon as we look at the coin, the property is defined and the superposition disappears.

The situation gets even more intriguing if we have two or more objects. In that case quantum mechanics predicts the existence of the type of superpositions known as entangled states. For example, we could have two coins (in two closed hands) which are in a superposition of two states — one in which both coins lie heads up, and one with both tails up. Because the coins are entangled, opening one hand defines the property of both coins, because if one lies heads up, the other must also be heads up. The reason is that having one coin heads up and the other tails up is not compatible with the states that create the superposition. The effect of observing one coin has an instantaneous consequence for the second — a strange phenomenon that is the basis of teleportation and other applications of entangled states.

In practice, we cannot do these experiments with coins, but physicists can use microscopic objects, such as atoms or photons. For an atom, for example, heads and tails equate to the quantum-mechanical property of spin — in other words, spin up and spin down. Up to four individual atoms have been entangled in this way in experiments carried out in Boulder⁶ and Paris⁷. Entangling more atoms would appear to be an even greater challenge because entangled states are easily destroyed. The reason for this is that the observations that destroy quantum superpositions do not necessarily require a human observer; they can result from any physical object (even the molecules of the air surrounding the object). So, to perform experiments with entangled atoms, the atoms must be isolated in such a way that

nothing can tell which state the superposition is in. The situation gets more difficult if the entangled objects are bigger and composed of more atoms, because then the interaction of a single atom with the environment may reveal the state of the superposition, destroying the entangled state.

How, then, have Julsgaard *et al.*¹ been able to entangle objects containing trillions of atoms? The trick is to have a superposition of two states: one in which slightly more than half of the atoms in each sample are spin up and another one in which slightly more than half are spin down. If the environment interacts with one (or more) atom and 'observes' that it is spin up, this is compatible with both states, so the superposition is not destroyed but slightly damaged. As a result, trillions of atoms must interact with the environment before the entanglement disappears.

The entangled state generated by Julsgaard *et al.* was created following a proposal made by Duan *et al.*⁸, in which a single light pulse interacts with the atoms of both samples and entangles them (Fig. 1). This is the first time two different atomic samples have been entangled in this way — using light — even though the samples are separated by some distance. Now that the experiment has been done, it should be relatively simple to entangle more than two atomic samples, or to teleport states of atomic samples. The work of Julsgaard and colleagues shows that in certain situations it may be more practical, and robust, to weakly entangle many atoms rather than strongly entangle a few of them. I expect that many new experiments will follow this one and help us better understand this strange theory. At the same time, they may well revolutionize the field of quantum information. ■

J. Ignacio Cirac is at the Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria.

e-mail: ignacio.cirac@uibk.ac.at

1. Julsgaard, B., Kozhkin, A. & Polzik, E. *Nature* **413**, 400–413 (2001).
2. Bennett, C. H. *et al.* *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
3. Bouwmeester, D. *et al.* *Nature* **390**, 575–579 (1997).
4. Boschi, D. *et al.* *Phys. Rev. Lett.* **80**, 1121–1124 (1998).
5. Furusawa, A. *et al.* *Science* **282**, 706–709 (1998).
6. Sackett, C. A. *et al.* *Nature* **404**, 256–259 (2000).
7. Rauschenbeutel, A. *et al.* *Science* **288**, 2024–2028 (2000).
8. Duan, L. M., Giedke, G., Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **84**, 2722–2725 (2000).

Daedalus

Ardent spirits

The fundamental laws of science, says Daedalus, must apply both to the material and spiritual worlds. These, he reckons, occupy the same space but are very weakly coupled, which is why ghosts are seldom observed and spiritual experience is rare too. The most fundamental scientific law is probably the 'zereth law of thermodynamics': that any two objects in contact reach thermal equilibrium. Over the age of the Universe, 18 gigayears or so, the material and spirit worlds must have equilibrated. The cosmic microwave background is 2.7 K; allowing for dark matter, the average temperature of the material world should be 10 or 15 K. The spirit world must have this temperature too, making it chilly indeed. Even the stars don't raise the average much.

Our nearest star is the Sun. The Michelson–Morley experiment found that the speed of light is the same in every solar season, making the Earth almost static with respect to the Sun. Spiritual experience and ghost observation are also independent of season, suggesting that the spirit world also moves slowly with respect to the Sun. This could be checked by flying a group of monks or nuns on Concorde or the Space Shuttle, looking to see whether spiritual insight is reduced as the velocity of the group rises with respect to the Sun.

All this leads up to a Daedalian theory of spiritual survival. To free itself for further adventures, a soul released from a body must try to warm itself, and will therefore head for the interior of the nearest star. A temperature of 16 million K will warm it usefully even with the weakest coupling, assuming that spiritual space moves slowly with respect to the solar interior. (By contrast, souls that 'go the wrong way' and head for the Earth's interior at a mere 6,000 K, will find themselves trapped in a local 'Hell'.) Furthermore, intelligent life on Earth over the years has released many souls, most of whom will have reached the Sun. They should perturb its internal reactions slightly, perhaps reducing its output of neutrinos, or setting them oscillating.

This novel theory of the solar neutrino deficit suggests a new way of seeking intelligent life around other stars. Rather than looking for the tiny librations caused by a circling planet, it would be easier to search directly for low neutrino emission by its primary star. Daedalus's recent directional neutrino telescope could prove vital.

David Jones

Obituary

Patrick D. Wall (1925–2001)

Major advances in biomedicine — such as the decoding of the human genome — increasingly depend on the application of ever more complex technologies, the harnessing of large teams of scientists and technicians, and the acquisition and management of huge budgets. Some of the most successful contemporary biomedical scientists are really like chief executive officers of large multinational corporations, more involved in managing and delegating than in experimenting or thinking. Patrick David Wall, who died on 8 August aged 76, was the antithesis of this kind of scientist.

Wall's major scientific contributions were twofold. First, he was integral in transforming the investigation of pain into a key area of modern neurobiology, one that is fully amenable to study at the molecular, cellular and systems levels. And second, he recognized that the ability of the nervous system to change — its plasticity — is fundamental to its function, both in normal circumstances and in disease.

Wall studied medicine at Oxford University during the Second World War, moving shortly afterwards to the United States. There, during brief sojourns at Yale, Chicago and Harvard, he mastered the art of electrophysiology and interacted with many of the major figures in the then new field of neuroscience. Afterwards, he settled for some years at the Massachusetts Institute of Technology. It was here that his fundamental work on pain and neuronal plasticity evolved, culminating in the seminal 'spinal gate control theory', which he developed with Ron Melzack and published in 1965. This theory basically proposes that there is a gating system in the central nervous system that determines when and how much pain we feel from an injury. It offers an explanation for why similar injuries can produce widely different sensations of pain, depending on whether a gate in the spinal cord opens or is blocked in response both to the pattern of input it receives and to influences from the brain.

In the late 1960s, Wall returned to Britain to lead the cerebral functions research group at University College London. It was here, and during his repeated visits to Israel, where he worked at the Hebrew University in Jerusalem, that he first argued and then showed that there are different types of pain. The pain that helps to protect us from tissue injury — for example, the sensation that we feel when we touch a hot pan that makes us



Neurobiologist who transformed the study of pain

move our hands away quickly — is quite different from the pain that arises when tissue is inflamed or nerves are damaged. Moreover, these two types of pain, the physiological and the pathological, involve different neurobiological mechanisms.

Before then, all pain was seen as the result of activation of a specific neuronal system — the 'pain pathway'. The intensity and duration of pain were thought to be determined only by the intensity and duration of the peripheral stimulus. But Wall recognized that pain is a sensory experience that shows a complex relationship to many different kinds of stimuli, some extrinsic and some intrinsic. A peripheral stimulus that would be expected to produce pain might not do so in some circumstances. Equally, a stimulus that would commonly be experienced as innocuous could, if nerves are damaged, elicit exquisite pain.

In studying this complex relationship, Wall investigated how damage to a peripheral nerve changes the excitability of nerve fibres — which begin to fire spontaneously — as well as the functional connections that they make. He also investigated the role of inhibitory systems in controlling the transfer of information in the nervous system. He recognized that a decrease in inhibition as well as increase in excitability of the nervous system is fundamental to many types of long-lasting pain.

In short, the unifying theme of much of Wall's work was his desire to decry the simple cartesian notion of a fixed pain system that connects the peripheral tissues

with the seat of conscious awareness in the brain. He also wanted to expose the complex events that lead to the varied feelings of discomfort and hurt that we call pain. Although he attacked many modern molecular approaches to the study of pain as forms of 'molecular phrenology' — what he saw as the ridiculous search for a pain gene — this approach is, ironically, beginning to validate his views by revealing the complexity of reactions to different types of injury. Pain involves change at many levels, from gene expression to activation of the brain's cortical region.

Wall also maintained that success in treating patients suffering from chronic pain requires an understanding of the underlying mechanisms, rather than trial-and-error therapies. One of his greatest strengths was that he recognized that pain is a personal experience that has to be studied in patients. He made important contributions to new approaches for managing pain, and actively encouraged the setting up of the hospice movement.

Wall's power as a scientist arose from his ideas, from his dedication to experiments — which he always conducted himself — and from his ability to communicate his thoughts in a way that enthralled, and sometimes appalled, all who listened to him or read his work. He enthralled because of his extraordinary capacity to translate dry scientific observations into a story that could be understood and wondered at. And he appalled because he was always willing to flout perceived wisdom and to expose, without any token of pseudocivility, what he saw as stupidity or charlatanism.

He also lacked the caution we usually expect of an experimentalist: he was always willing to use his data as a platform to develop his ideas. Yet these ideas were well thought out, and based on a deep understanding of neurobiology, uncanny intuition, and an insight into what the data mean in a larger context. This capacity to think deeply about his findings, rather than just catalogue them, meant that his interpretations always seemed to be ahead of the data. This enraged some of his competitors, but Pat saw that as a compliment. He was both an extraordinary scientist and a forceful individual, whose powerful political convictions and personal eccentricities made him one of a kind.

Clifford J. Woolf

Clifford J. Woolf is in the Neural Plasticity Research Group, Department of Anesthesia and Critical Care, Massachusetts General Hospital and Harvard Medical School, 149 13th Street, Charlestown, Massachusetts 02129, USA. e-mail: woolf.clifford@mgh.harvard.edu

Transatlantic robot-assisted telesurgery

ATM technology now enables operations to be performed over huge distances.

The introduction of robotic and computer technology into surgical operations allows dexterity to be increased^{1–3} and surgical procedures to be carried out from a distance (telesurgery)⁴. But until now, the distance feasible for remote telesurgery was considered to be limited to a few hundred miles¹ by the time lag of existing telecommunication lines. Here we show that robot-assisted remote telesurgery can be safely carried out across transoceanic distances. The ability to perform complex surgical manipulations from remote locations will eliminate geographical constraints and make surgical expertise available throughout the world, improving patient treatment and surgical training.

Telementoring and limited surgical assistance over long distances have been reported^{5–10}, but technical limitations have so far prevented the performance of a complete operation. Telesurgery was expected to involve only telementoring, rather than remote manipulation, for the foreseeable future¹¹. Limiting factors have been the time for converting video images and surgical movements into electronic signals, and the bandwidth and time delay of existing telecommunication lines^{12–14}.

To investigate the impact of variable time delays on surgical manipulations, we carried out robot-assisted laparoscopic cholecystectomy (gall-bladder removal) on a pig, transmitting the signals from the surgeon's console in Strasbourg, France, to Paris and back — a total distance of about 1,000 km. The time lag was artificially increased from 20 milliseconds (standard time delay) up to 551.5 ms. The limit of the acceptable time delay in terms of a surgeon's perception of safety was roughly 330 ms.

To measure transoceanic latencies and verify the feasibility of an intercontinental surgical operation, we attempted robot-assisted laparoscopic cholecystectomy in a porcine model across a round-trip distance of over 14,000 km. The operator site was located in New York and the animals were in Strasbourg.

The robotic system (ZEUS, Computer Motion, California) consisted of two separate subsystems ('surgeon-side' (Fig. 1) and 'patient-side' (Fig. 2)). The two sites were connected through a high-speed terrestrial optical-fibre network (FranceTelecom/Equant) that transports data through dedicated connections using asynchronous transfer mode (ATM) technology. A bandwidth of 10 megabits per second has been reserved through a



Figure 1 Surgeon operating the robotic console in New York to remove the gall bladder of a patient in France.

network that interconnects applications at both sites using a network termination unit (NTU), which provides a multiservice path to different applications.

To monitor and measure its level of quality, the NTU sender inserted 'operating and maintenance' packets within user data flow, which were extracted and analysed by the remote NTU receiver. By analysing these packets and comparing the number of user packets initially sent to those that were actually received, we determined the number of lost packets.

Operation and maintenance tools revealed that no ATM packet was lost during any surgical procedure. The round-trip delay by ATM transport was 78–80 ms. Adding 70 ms for video coding and decoding, plus a few milliseconds for rate adaptation and Ethernet-to-ATM packet conversion, movements executed by the surgeon in New York were apparent within 155 ms on his video screen.

We successfully carried out laparoscopic remote robotic cholecystectomy in six pigs. At the operative site in New York, two surgeons carried out the procedure and a third monitored the recording. Two surgeons at the remote site set up the trocar position and were able to inactivate the robot if necessary for safety reasons. Coagulation was coordinated by voice command between the sites. The mean time for cholecystectomy was 45 min (range, 26–78 min).

The three surgeons in New York subjectively evaluated, in a blinded manner, the quality of the image, the impact of time lag on performance, the coordination and

safety of the use of electrocautery, and the overall safety of the procedure. Evaluation was on a 0–10 scale (where 0 is the worst possible and 10 the best possible). Scores were 9.1 for the quality of image, 8.5 for the impact of time lag (0, unacceptable impact; 10, imperceptible impact), 9.2 for coordination of electrocautery, and 8.7 for overall safety. These scores reflect the high confidence of the surgeons in controlling the surgical movements as well as the general reliability of the system.

Note added in proof: We have successfully carried out a remote laparoscopic cholecystectomy in a 68-year-old female after obtaining ethical committee approval and informed consent from the patient. The surgeons were in New York and the patient was in Strasbourg. The circuit used was the same as the one described here; the mean total time delay was 155 ms. The time taken to set up the robotic system was



Figure 2 Robotic arms at the remote site in Strasbourg.

16 min, and the gallbladder was dissected in 54 min, without any intra-operative complication. The post-operative period was uneventful and the patient was discharged 48 h later.

Jacques Marescaux*, **Joel Leroy***,
Michel Gagner†, **Francesco Rubino***,
Didier Mutter*, **Michel Vix***,
Steven E. Butner‡, **Michelle K. Smith***

*IRCAD European Institute of Telesurgery, Louis Pasteur University, 1 Place de l'Hopital, 6700 Strasbourg, France

e-mail: jacques.marescaux@ircad.u-strasbg.fr

†IRCAD European Institute of Telesurgery, and Department of Laparoscopic Surgery, Mount Sinai Medical Center, New York 10029, USA

‡Department of Electrical and Computer Engineering, University of California,

Santa Barbara, California 93106, USA

1. Satava, R. M. *Arch. Surg.* **134**, 1197–1202 (1999).
2. Haluck, R. S. & Krummel, T. M. *Arch. Surg.* **135**, 786–792 (2000).
3. Buckingham, R. A. & Buckingham, R. O. *Br. Med. J.* **311**, 1479–1482 (1995).
4. Link, R. E., Schulam, P. G. & Kavoussi, L. R. *Urol. Clin. N. Am.* **28**, 177–188 (2001).
5. Janetschek, G., Bartsch, G. & Kavoussi, L. R. *J. Urol.* **160**, 1413–1416 (1998).
6. Cubano, M. et al. *Surg. Endosc.* **13**, 673–678 (1999).
7. Rosser, J. C. Jr et al. *J. Am. Coll. Surg.* **189**, 397–404 (1999).
8. Lee, B. R. et al. *World J. Urol.* **16**, 367–370 (1998).
9. Micali, S. et al. *J. Endourol.* **14**, 493–496 (2000).
10. Cheah, W. K., Lee, B., Lenzi, J. E. & Goh, P. M. Y. *Surg. Endosc.* **14**, 1085 (2000).
11. Mack, M. J. *J. Am. Med. Assoc.* **285**, 568–572 (2001).
12. Fabrizio, M. D. et al. *J. Endourol.* **14**, 133–138 (2000).
13. Satava, R. M. & Jones, S. B. in *Cybersurgery: Advanced Technologies for Surgical Practice* (ed. Satava, R. M.) 141–154 (Wiley-Liss, New York, 1998).
14. Broderick, T. J. et al. *Telemed. J. e-Health* **7**, 47–53 (2001).

Microbiology

Cyanobacteria track water in desert soils

Cyanobacteria develop as large, cryptic populations in the topsoil of arid land, where plant cover is restricted, water is scarce and harsh microenvironmental conditions prevail. Here we show that some cyanobacteria can actively move in response to wetting or drying events by migrating to the soil surface or retreating to their refuge

below. This ability to follow water, which to our knowledge has not been demonstrated before in microbes, may turn out to be important for microbial terrestrial populations in general.

Although cells direct their movements in response to environmental stimuli^{1,2}, so far no microorganism has been shown to exhibit a tactic response to water, perhaps the most crucial molecule in biology. We tested the response to soil wetting of cyanobacteria — important primary producers in topsoils that support only sparse

plant growth, notably in hot and cold deserts³ — and found that they move away from surface-desiccation fronts. This behaviour, together with a tactic response to light, results in vertical migration patterns that probably optimize the position of these microorganisms in the soil matrix for incoming drought periods, thereby contributing to their success in harsh habitats.

Using microscale measurements of surface spectral reflectance⁴, we monitored the vertical migration of cyanobacteria in samples of arid soil from badlands in Spain. This soil, in its original desiccated state, contained a subsurface population of a filamentous oscillatorian cyanobacterium (*Oscillatoria* sp.; estimated population maximum at a depth of about 2 mm).

Greening of the surface was evident within minutes of wetting (Fig. 1a) and increased linearly for up to 2 hours. Microscopic observation of wetted samples showed that this greening was due to the appearance of cyanobacterial filaments at the surface (Fig. 1b, c). The greening could be slowed or prevented by exposing the soil to high-intensity light, indicating the involvement of cellular photoresponses, but also occurred in the dark (data not shown), as found previously for cyanobacteria in other habitats⁵.

When the soils were allowed to dry under unchanged illumination, the greening stopped and eventually reversed, with the cyanobacteria apparently moving downwards. This reversal also took place in complete darkness. Upward and downward migrations could be induced repeatedly (Fig. 1d) by wetting and drying cycles. We conclude that the greening reversal is not due to a photoresponse or to a chemoreponse to oxygen, pH or CO₂ gradients⁶, which are opposite under dark and light conditions⁷.

When inhibitors of cellular ATP generation were added to the soil (Fig. 1e), the greening reversal was completely inhibited, indicating that cellular energy is required to sustain gliding motility. The cyanobacterial retreat is therefore unlikely to be a purely physical effect driven by surface tension at the air–water interface. Our evidence indicates that this cyanobacterium can either sense changes in water availability and respond with tactic behaviour, or can hold itself in a wet environment.

In this cyanobacterium, this behaviour overwhelmed the tendency to seek optimal illumination, which would otherwise keep the population at the soil surface. The cyanobacteria are thus able to reach new refuges and are less likely to be trapped in a desiccated, near-surface environment. This serves to reduce population photodamage and losses through erosion, and to increase survival during drought periods.

Cells may sense short-term changes in external water availability, perhaps through mechanosensitive channels⁸, which may

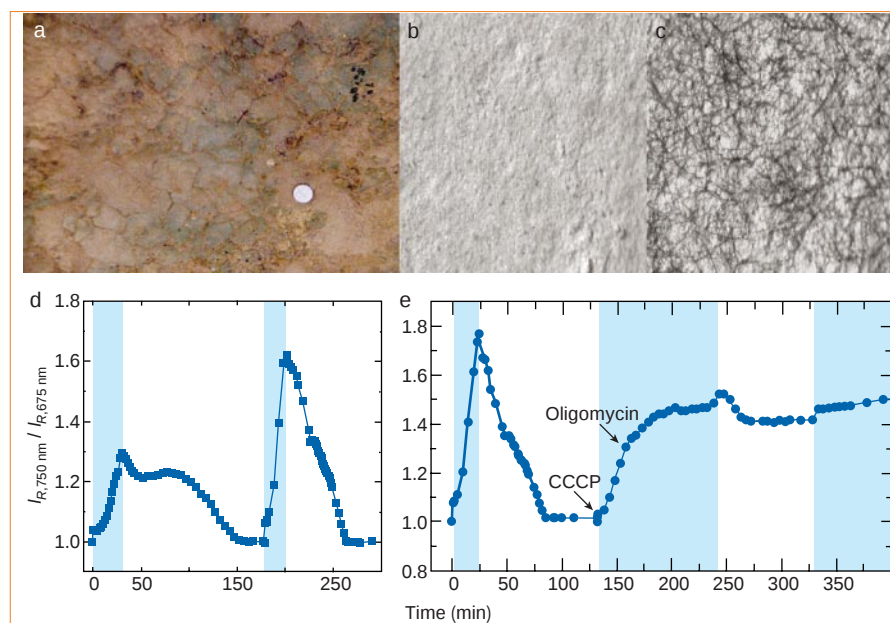


Figure 1 Vertical migrations of arid-soil cyanobacteria in response to wetting and drying. **a**, Greening of soil from Bardena Blanca (Navarre, Spain) after artificial watering. Green patches are populations of oscillatorian cyanobacteria; black dots are colonies of the lichen *Collema*. A coin (lower right; diameter, 17.5 mm) is shown for scale. **b**, **c**, High-magnification photographs of soil surface before (**b**) and 20 min after (**c**) wetting (exposure and brightness were modified for optimal clarity). Trichomes (filaments) of oscillatorian cyanobacteria are visible in **c** on the soil surface. **d**, **e**, Dynamics of greening and its reversal, as determined from the ratio of intensity of infrared to red reflectance at the wavelengths indicated. Shaded areas represent greening-response periods under flooding. In **d**, a dry field sample is wetted at time zero: greening is measurable within minutes; at 30 min, excess water is retrieved, drying and greening reversal begin, causing the ratio to return to its original value after 150 min. The cycle is repeated between 150 and 300 min. In **e**, the effect of sequential addition of inhibitors (arrows) of oxidative phosphorylation (carbonyl cyanide chlorophenylhydrazone, CCCP) and of cellular ATP generation (oligomycin) on the cycle of greening and reversal is shown for the second cycle.

indicate that the cyanobacterial response is hydrotactic. Alternatively, cells may rely on sensing parameters that vary with desiccation, such as salinity, temperature or the presence of other chemical species⁹.

We have also seen this behaviour in other desert-soil cyanobacteria — for example, the common *Microcoleus vaginatus*^{3,10}. If it is widespread among prokaryotes, this response may be crucial for survival and dispersal in terrestrial environments, as most bacteria are not bound to the surface by phototaxis. Soil and deep subsurface bacteria that exist in stationary populations in dry soil may not only be passively transported by groundwater but may also actively follow a moving water pocket in order to exploit it.

Ferran Garcia-Pichel*, Olivier Pringault†

*Department of Microbiology, Arizona State

University, Tempe, Arizona 85287, USA

e-mail: ferran@asu.edu

†Laboratoire d'Océanographie Biologique,

Université Bordeaux 1, CNRS-UMR-EPOC 5805,

2 Rue du Prof Jolyet, 33120 Arcachon, France

1. Armitage, J. *Adv. Microb. Physiol.* **41**, 229–249 (2000).
2. Wilkinson, P. C., Komaikoma, M. & Newman, L. *Autoimmunity* **26**, 55–72 (1997).
3. Garcia-Pichel, F., López-Cortés, A. & Nübel, U. *Appl. Environ. Microbiol.* **67**, 1902–1910 (2001).
4. Bebout, B. M. & Garcia-Pichel, F. *Appl. Environ. Microbiol.* **61**, 4215–4222 (1995).
5. Garcia-Pichel, F., Mechling, M. & Castenholz, R. W. *Appl. Environ. Microbiol.* **60**, 1500–1511 (1994).
6. Malin, G. & Walsby, A. E. *J. Gen. Microbiol.* **131**, 2643–2652 (1985).
7. Garcia-Pichel, F. & Belnap, J. J. *Phycol.* **32**, 774–782 (1996).
8. Martinac, B., Adler, J. & Kung, C. *Nature* **348**, 261–263 (1990).
9. Li, C., Boileau, J. A., Kung, C. & Adler, J. *Proc. Natl Acad. Sci. USA* **85**, 9451–9455 (1988).
10. Belnap, J. & Gardner, J. S. *Great Basin Naturalist* **53**, 40–47 (1993).

Intracellular signalling

Receptor-specific messenger oscillations

The cytosolic molecule inositol-1,4,5-trisphosphate (InsP₃) acts as a messenger to link receptors at the cell surface with alterations in calcium concentration inside the cell¹, but it is not clear how production of InsP₃ is related to the often-complex calcium response². Here we use a fluorescent biosensor to visualize InsP₃ synthesis in individual cells in real time and show that this is periodically switched on and off in a receptor-specific manner. Our findings are consistent with intracellular calcium oscillations being generated by either fluctuating or sustained concentrations of InsP₃, which may allow diversity of signalling through the same signal-transduction pathway.

Oscillations in intracellular calcium concentration ([Ca²⁺]_i) induced by G-protein-coupled receptors in the membrane provide a versatile encoding mechanism that uses

variations in the amplitude, frequency and duration of signals to control cellular processes^{1,3,4}. Models to explain these oscillations are broadly based on dynamic uncoupling of the phospholipase C (PLC)/InsP₃ signalling pathway, or on the self-propagating regulatory properties of Ca²⁺ on the InsP₃ receptor (known as Ca²⁺-induced Ca²⁺ release)^{2,5–7}. Distinction between the two schemes relies on whether InsP₃ oscillations, from repetitively switching phospholipase C on and off, are the driving force, or whether Ca²⁺ alone controls this process by enhancing or inhibiting its own release from internal stores at low and high concentrations, respectively.

The pleckstrin-homology (PH) domain of the PLC-δ1 enzyme allows these alternatives to be tested, as it binds to the precursor of InsP₃, phosphatidylinositol-4,5-bisphosphate, which is associated with the cell membrane, but translocates to the cytosol after activation of phospholipase C and InsP₃ synthesis^{8–10}. Tagging the PH domain with green fluorescent protein (GFP) allows this process to be visualized (Fig. 1a) and the fractional increase in cytosol fluorescence can be used as an index of InsP₃ production (Fig. 1b–d).

We studied two types of PLC-linked G-protein-coupled receptor that induce oscillatory Ca²⁺ responses: the metabotropic glutamate receptor, mGluR5a, and the M₃-muscarinic receptor. Stimulation through mGluR5a produced oscillatory patterns of InsP₃ production, which were consistently paralleled by changes in [Ca²⁺]_i (Fig. 1a–c). We found that oscillations were controlled by protein kinase C, as prolonged treatment with phorbol ester to downregulate this enzyme resulted in sustained InsP₃ production, which failed to oscillate even at low agonist concentrations (Fig. 1b, inset).

These results reveal, to our knowledge for the first time, [Ca²⁺]_i changes that are induced by InsP₃ oscillations through dynamic and rapid uncoupling. This probably occurs through phosphorylation by protein kinase C of a single residue in mGluR5a (ref. 11). In contrast, although oscillatory and sustained Ca²⁺ signals induced by M₃-muscarinic receptors are dependent on agonist concentration (Fig. 1d), we detected no InsP₃ oscillation. This suggests that [Ca²⁺]_i alone does not significantly stimulate InsP₃ production and that mGluR5a-induced InsP₃ oscillations are not secondary to increases in [Ca²⁺]_i.

We propose that at least two mechanisms — Ca²⁺-induced Ca²⁺ release and dynamic uncoupling determined by the intrinsic properties of the receptor — can drive [Ca²⁺]_i oscillations in the same cell background. For the latter scheme, which is typified by mGluR5a, cyclical desensitization and resensitization dynamically controls the activity of phospholipase C. In contrast,

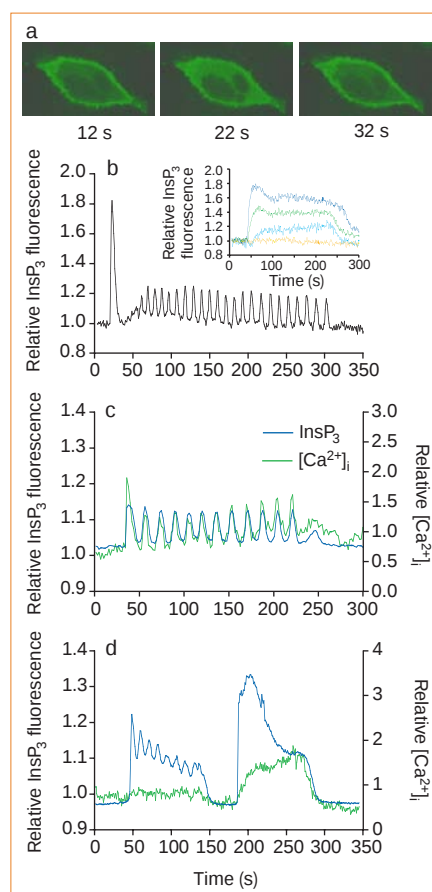


Figure 1 Single-cell imaging of intracellular InsP₃ and [Ca²⁺]_i in Chinese hamster ovary cells expressing mGluR5a or M₃-muscarinic receptors. **a**, Expression of the fusion protein GFP-PH_{PLCδ1} in an mGluR5a-expressing cell before (12 s) and during (22 and 32 s) treatment with 1 mM glutamate, showing transient translocation of the GFP fusion protein to the cytosol at 22 s. Times correspond to those for the trace in **b**, where InsP₃ oscillations are represented as the fractional change in cytosolic fluorescence relative to the basal value. Inset, InsP₃ production is sustained after downregulation of protein kinase C by overnight incubation with 1 μM phorbol dibutyrate (glutamate concentrations: blue, 100 μM; green, 10 μM; cyan, 3 μM; yellow, 1 μM). **c**, Oscillations in InsP₃ are closely paralleled by similar changes in [Ca²⁺]_i in an mGluR5a-expressing cell treated with 10 μM quisqualate. Intracellular Ca²⁺ concentration was determined using Fura-2 dye and is expressed as the ratio of fluorescence at 340 nm and 380 nm. **d**, Traces showing the changes in InsP₃ and [Ca²⁺]_i in a single M₃-expressing cell treated with 0.1 μM methacholine, washed and treated again with 1 μM methacholine.

muscarinic-receptor-induced Ca²⁺ oscillations probably result from classic Ca²⁺-induced Ca²⁺ release after an initial stimulus of a threshold InsP₃ concentration, although participation by other messengers such as cyclic ADP-ribose cannot be ruled out¹².

This finding has important implications for cell signalling. Dynamic uncoupling allows oscillatory Ca²⁺ responses to occur even at maximal agonist concentration, and each Ca²⁺ oscillation is thus under the temporal control of the InsP₃ signal. This links the encoded Ca²⁺ signal with concomitant phosphorylation reactions that take place in the diacylglycerol/protein kinase C arm of

the bifurcating phospholipase C pathway.

In contrast, Ca^{2+} -induced Ca^{2+} oscillations at low agonist concentration may partially isolate the activities of Ca^{2+} release from those of protein kinase C activation. The full range of oscillatory and sustained-plateau Ca^{2+} signals can thus be elicited from a single receptor. This complex Ca^{2+} signalling not only increases the flexibility with which cellular processes can be controlled, but also allows different G-protein-coupled receptors, using the same transmembrane signalling in the same cell, to achieve specificity.

**Mark S. Nash, Kenneth W. Young,
R. A. John Challiss, Stefan R. Nahorski**

Department of Cell Physiology and Pharmacology,
Medical Sciences Building, University of Leicester,
University Road, Leicester LE1 9HN, UK

e-mail: srn@le.ac.uk

1. Berridge, M. J., Lipp, P. & Bootman, M. D. *Nature Rev. Mol. Cell Biol.* **1**, 11–21 (2000).
2. Taylor, C. W. & Thorn, P. *Curr. Biol.* **11**, R352–R355 (2001).
3. De Koninck, P. & Schulman, H. *Science* **279**, 227–230 (1998).
4. Dolmetsch, R. E., Xu, K. & Lewis, R. S. *Nature* **392**, 933–936 (1998).
5. Harootyanian, A. T., Kao, J. P., Paranjape, S. & Tsien, R. Y. *Science* **251**, 75–78 (1991).
6. Bird, G. S., Rossier, M. F., Obie, J. F. & Putney, J. W. Jr *J. Biol. Chem.* **268**, 8425–8428 (1993).
7. Thomas, A. P., Bird, G. S., Hajnóczky, G., Robb-Gaspers, L. D. & Putney, J. W. *FASEB J.* **10**, 1505–1517 (1996).
8. Hirose, K., Kadowaki, S., Tanabe, M., Takeshima, H. & Iino, M. *Science* **284**, 1527–1530 (1999).
9. Nash, M. S., Young, K. W., Willars, G. B., Challiss, R. A. & Nahorski, S. R. *Biochem. J.* **356**, 137–142 (2001).
10. van Der Wal, J., Habets, R., Varnai, P., Balla, T. & Jalink, K. *J. Biol. Chem.* **276**, 15337–15344 (2001).
11. Kawabata, S. et al. *Nature* **383**, 89–92 (1996).
12. Petersen, O. H. & Cancela, J. M. *Trends Neurosci.* **22**, 488–494 (1999).

COMMUNICATION ARISING

Hospital waiting-lists

Do power laws imply self-regulation?

Negative feedback leading to self-regulatory behaviour is an important phenomenon that affects time-series fluctuations in a range of systems and is critical in forecasting and management, particularly when complex dynamics are possible. Smethurst and Williams¹ argue that the lengths of waiting-lists to see hospital consultants are self-regulating, on the grounds that the relative changes in the size of waiting-lists follow a power law, with large changes being relatively rare compared with small ones. Here we show that similar power laws can also be obtained from unregulated, random time series. The existence of a power law that governs fluctuations in time series is not sufficient to prove the existence of self-regulatory behaviour, and we argue that a more sophisticated analysis is required.

Figure 1a shows a power law derived

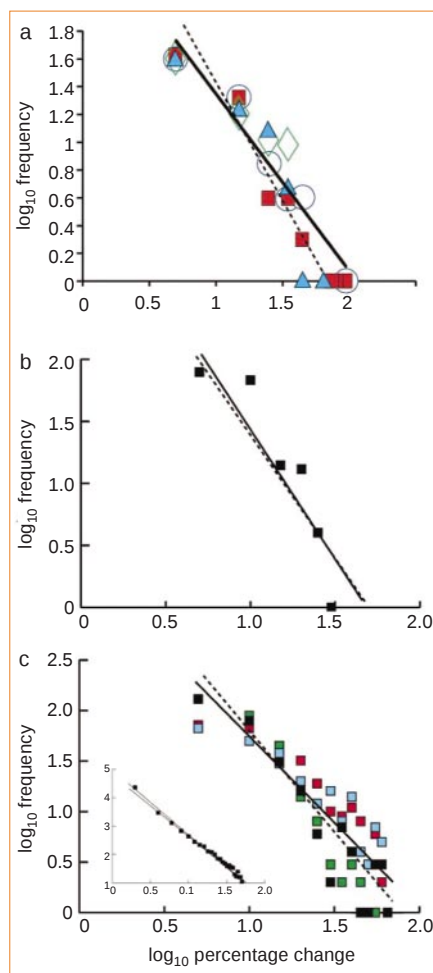


Figure 1 Power-law behaviour in time series. **a–c**, Frequency distributions of relative changes in numbers (measured as $(N_{t+1} - N_t)/N_t$, where N is the number at current time t , or at one time period later, $t+1$) plotted on a double-logarithmic scale. **a**, Frequency distribution of the extent of length changes in hospital waiting-lists for consecutive months (different symbols signify different consultants); data from ref. 1. The frequency of large variations declines according to a power law (solid line, regression). **b**, Similar analysis of changes in the number of people at the bar of the Butcher's Arms public house in Headington, near Oxford. **c**, Similar analysis of data from discrete-time random walks. Results of random walks of length 300 time units are shown; different symbols denote separate replicates. Inset, asymptotic distribution, from a series of length 30,000 time units. These random walks (brownian motion) assume changes in numbers that are normally distributed (with zero mean); at any time t , the change in number is independent of the number at that time. Hence, this system shows no self-regulatory behaviour — in terms of a waiting-list or queue, this would imply that people join or leave at random, with no regard for the number of other people waiting. In **a–c**, dashed lines have a slope of -2 , the expected value for a log-normal distribution of relative fluctuations.

from the data of Smethurst and Williams¹. In their analysis, the rarity of large changes in the length of waiting-lists is used as evidence for negative feedback: people are more likely to join short rather than long queues, thereby buffering changes. Figure 1b reveals a similar relationship in the number of people waiting at the bar of a public house. The similarity of the plots for the two

systems could be taken as evidence of a general power law governing the behaviour of queues. However, the data in Fig. 1a, b have not been tested against a null model, and it is not clear what pattern would be expected in a system with no self-regulatory behaviour.

The appropriate null model for a system with no regulation is a random walk (brownian motion). If data are following a random walk (that is, if the time series is unregulated), the frequency distribution of relative changes will be expected to show power-law-like behaviour. This is because even in unregulated time series, relatively large fluctuations would be expected to be less common than relatively small ones.

To illustrate this point, Fig. 1c shows the results of applying Smethurst and Williams' analysis to randomly generated data from brownian-motion time series. These time series yield frequency distributions of relative changes that are strikingly similar to the plots in Fig. 1a, b, and which have a power-law slope that is indistinguishable from -2 . This indicates that Smethurst and Williams' data are entirely consistent with models in which self-regulation does not occur. The existence of this power-law relationship cannot therefore be used to infer the operation of self-regulatory mechanisms.

Providing evidence for such negative feedback is difficult, as shown, for example, by the problems of detecting density-dependence in ecology^{2,3}. We suggest three approaches to the problem of detecting negative feedback in hospital waiting-lists, on the basis of approaches taken by population ecologists. One is to adjust queue lengths experimentally and examine the responses². The second is to plot the rate of change in queue length against the length of the queue, although this is subject to statistical problems^{2,3}. The third is to use behaviour-based population modelling⁴ to determine the range of trade-offs made by individuals — for example, specific trade-offs can be quantified between the expected waiting time for treatment and the probability of choosing private treatment. Knowledge of such responses could then be applied to reducing waiting times. By contrast, relationships of the type shown in Fig. 1 constitute a poor method of characterizing and forecasting such systems.

**Robert P. Freckleton*,
William J. Sutherland†**

*Department of Zoology, University of Oxford,
Oxford OX1 3PS, UK

e-mail: robert.freckleton@zoology.oxford.ac.uk

†School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

1. Smethurst, D. P. & Williams, H. C. *Nature* **410**, 652–653 (2001).
2. Bulmer, M. *Biometrics* **31**, 901–911 (1975).
3. Rothery, P. in *Insect Population* (eds Dempster, J. P. & McLean, I. F. G.) 32–68 (Chapman & Hall, London, 1998).
4. Sutherland, W. J. *From Individual Behaviour to Population Ecology* (Oxford Univ. Press, Oxford, 1996).

Two-component circuitry in *Arabidopsis* cytokinin signal transduction

Ildoo Hwang & Jen Sheen

Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA

Cytokinins are essential plant hormones that are involved in shoot meristem and leaf formation, cell division, chloroplast biogenesis and senescence. Although hybrid histidine protein kinases have been implicated in cytokinin perception in *Arabidopsis*, the action of histidine protein kinase receptors and the downstream signalling pathway has not been elucidated to date. Here we identify a eukaryotic two-component signalling circuit that initiates cytokinin signalling through distinct hybrid histidine protein kinase activities at the plasma membrane. Histidine phosphotransmitters act as signalling shuttles between the cytoplasm and nucleus in a cytokinin-dependent manner. The short signalling circuit reaches the nuclear target genes by enabling nuclear response regulators ARR1, ARR2 and ARR10 as transcription activators. The cytokinin-inducible *ARR4*, *ARR5*, *ARR6* and *ARR7* genes encode transcription repressors that mediate a negative feedback loop in cytokinin signalling. Ectopic expression in transgenic *Arabidopsis* of *ARR2*, the rate-limiting factor in the response to cytokinin, is sufficient to mimic cytokinin in promoting shoot meristem proliferation and leaf differentiation, and in delaying leaf senescence.

Cytokinins have long been recognized as essential plant hormones that are involved in diverse processes of plant growth and development. These processes include cell division, shoot initiation, leaf and root differentiation, chloroplast biogenesis and senescence^{1,2}. The cellular, molecular and biochemical mechanisms underlying the actions of cytokinins have not been elucidated. However, recent genetic and molecular studies in various plant species have suggested the involvement of two-component signalling proteins in cytokinin signal transduction^{3–7}.

Two-component systems—consisting of a histidine protein kinase that senses the input and a response regulator that mediates the output—control signal transduction pathways in many prokaryotes and in some eukaryotes. The signalling pathway is initiated when a histidine protein kinase sensor, modulated by an environmental cue, phosphorylates its own conserved histidine residue and transfers the phosphate to a conserved aspartate residue of a response regulator. In some cases, additional phosphotransfer steps may intervene between the histidine protein kinase and response regulator, mediated by a histidine phosphotransmitter^{8–10}. The completion of the *Arabidopsis* genome sequence has revealed over 40 genes encoding putative AHK, AHP and ARR proteins, suggesting an important involvement of the ancient and conserved signalling mechanism in many facets of plant cell regulation^{11,12}. The identification of conserved histidine protein kinase signature motifs in the photoreceptor phytochrome, a putative osmosensor, and the ethylene and cytokinin receptors of *Arabidopsis* further supports this view^{6,13–15}.

In eukaryotes, a two-component circuit often provides a link between the sensing of an environmental cue and a MAP kinase (MAPK) signalling cascade. For example, the SLN1/YPD1/SSK1 phosphorelay in yeast translates a change in osmolarity into the phosphorylation state of the response regulator SSK1, which then modulates the HOG1 MAPK cascade in the cytosol to control gene expression^{16–18}. Plant histidine protein kinases, such as the ethylene receptor (ETR1) and a putative osmosensor (AHK1), may also use phosphorelays to transmit signals to MAPK cascades^{14,15,19,20}. These eukaryotic sensor kinases seem to have negative roles as their inactivation results in downstream signalling. However, the implication of the putative AHK proteins, CKI1 and CRE1 (refs 3, 6, 21), in cytokinin signalling raises the possibility that a two-component system may activate cytokinin signal transduction in *Arabidopsis*. The lack of a physiological and genetically manipulable plant cell

system has made it difficult to discern the mechanism by which a histidine protein kinase may mediate cytokinin signalling or any other plant response.

We have developed a transient expression assay on the basis of the transcriptional activation of a cytokinin primary response gene, *ARR6*, using mesophyll protoplasts of *Arabidopsis* leaves. This assay has enabled us to perform functional genomic analysis of the two-component regulators, and to decipher a cytokinin signalling pathway in *Arabidopsis*. The pathway does not follow the established eukaryotic histidine protein kinase and MAPK cascade model, but rather integrates multiple histidine protein kinase activities to common AHP proteins, which serve as cytoplasm/nuclear shuttles, and to distinct ARR proteins in the nucleus. The *Arabidopsis* cytokinin signal transduction pathway consists of four principal steps: histidine protein kinase sensing and signalling; AHP translocation; ARR-dependent transcription activation; and a negative feedback loop through cytokinin-inducible ARR genes. Analyses of transgenic tissues and plants support the importance of this central signalling pathway in diverse cytokinin responses.

Cytokinin-inducible transcription in leaf cells

To elucidate the regulatory circuitry in cytokinin signal transduction, we developed a leaf cell assay on the basis of cytokinin-inducible transcription in *Arabidopsis* mesophyll protoplasts. The 2.4-kilobase (kb) promoter of an *Arabidopsis* cytokinin primary response gene—encoding the response regulator 6 (*ARR6*)—was fused to the firefly luciferase (*LUC*) coding sequence to create a reporter construct, *ARR6-LUC*. In transfected protoplasts, the activity of *ARR6-LUC* was specifically induced by cytokinin but not by other plant hormones such as abscisic acid (ABA) or auxin (Fig. 1a). In the same system, the *GH3* promoter was activated only by auxin, whereas the *RD29A* promoter was induced only by abscisic acid (ABA), demonstrating the specificity of three plant hormone responses in protoplasts (Fig. 1a). The activity of *ARR6-LUC* was induced by physiological concentration of a natural cytokinin *trans*-zeatin (*t*-zeatin) from 1 to 100 nM (Fig. 1b). To further show the specificity of the cytokinin response, various active and inactive cytokinin analogues were examined. Only active cytokinins, *t*-zeatin, 2-isopentenyladenine (2-IP) and 6-benzyladenine (BA), induced the reporter gene *ARR6-LUC* (Fig. 1c). The response in the protoplast system is similar to the cytokinin activation of various ARR genes shown in plants^{4,5}. Thus, we have

established a reliable and physiological system to dissect the regulatory components in the cytokinin signal transduction pathway in *Arabidopsis*.

Cytokinin signalling by distinct AHK receptors

CKI1, a hybrid histidine protein kinase with a conserved receiver domain, has been implicated in cytokinin responses³. The enhanced CKI1 expression in *Arabidopsis shooty callus* mutants also supports its role in cytokinin signalling²². However, the biochemical mechanism of CKI1 function and its role in cytokinin signalling are still unknown. To determine whether CKI1 mediates cytokinin responsive transcription, the full-length *CKI1* gene was cotransfected with

ARR6-LUC into *Arabidopsis* protoplasts. Notably, wild-type CKI1 activated the *ARR6* promoter without exogenous *t*-zeatin (Fig. 1d). It is possible that overexpression of CKI1 renders protoplasts hypersensitive to endogenous cytokinin and/or exceeds the capacity of negative regulators. Alternatively, CKI1 could encode a constitutively active histidine protein kinase connected to the cytokinin signal transduction pathway. Cytokinin treatment did not significantly enhance the reporter gene activity in the presence of CKI1 (Fig. 1d). To see whether the histidine protein kinase activity and phosphoryl transfer is required for CKI1 activation of the *ARR6* promoter, the conserved His405 and Asp1050 residues in the histidine protein kinase and receiver domains, respectively, were

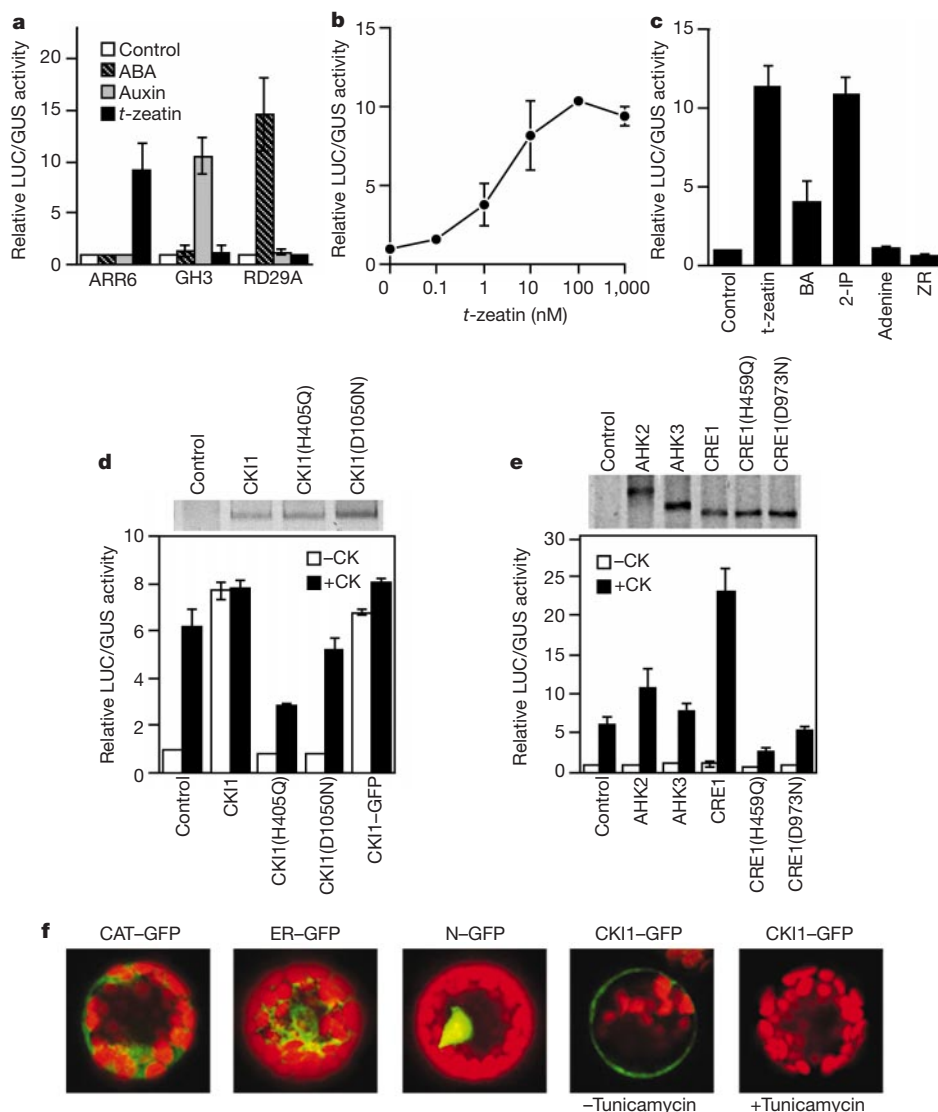


Figure 1 Cytokinin signalling is initiated by multiple histidine protein kinase receptors. **a**, Specificity of plant hormone responses in the *Arabidopsis* mesophyll protoplast transient expression system. Protoplasts were transfected with *UBQ10-GUS* (internal control) and *ARR6-LUC* (*ARR6*), *GH3-LUC* (*GH3*) or *RD29A-LUC* (*RD29A*) plasmid DNA. Transfected protoplasts were incubated without (control) or with 100 nM *t*-zeatin, 1 μ M IAA, or 100 μ M ABA. **b**, Cytokinin dose response for the *ARR6* promoter induction. **c**, Induction of the *ARR6* promoter by different cytokinins. All chemicals are at 100 nM. BA, 6-benzyladenine; 2-IP, 2-isopentenyladenine; ZR, zeatin riboside. **d**, CKI1 activation of *ARR6-LUC* requires histidine protein kinase activity and phosphoryl transfer. Protoplasts were cotransfected with the *ARR6-LUC* reporter and an effector plasmid expressing CKI1, CKI1(H405Q), CKI1(D1040N) or CKI1-GFP. Vector DNA was used as a control. The top panel shows the ³⁵S-methionine labelled CKI1, CKI1(H405Q) and

CKI1(D1050N) proteins after immunoprecipitation. Transfected protoplasts were treated without (–CK) or with (+CK) 100 nM *t*-zeatin. **e**, CRE1 confers cytokinin hypersensitivity on the activation of the *ARR6-LUC* reporter. Protoplasts were cotransfected with the *ARR6-LUC* reporter and an effector plasmid expressing AHK2, AHK3, CRE1(AHK4/WOL), CRE1(H459Q) or CRE1(D973N). The top panel shows the expression of AHK2, AHK3, CRE1(AHK4/WOL), CRE1(H459Q) and CRE1(D973N) proteins. **f**, CKI1 is localized at the plasma membrane. Protoplasts were transfected with CKI1-GFP or various GFP marker plasmid DNA. Tunicamycin (1 μ g ml^{–1}) treatment was for 12 h. The subcellular GFP markers are CAT-GFP for the cytosol, ER-GFP⁴⁷ for the endoplasmic reticulum, and N-GFP (DOF-GFP)⁴⁸ for the nucleus. The red fluorescence is from chlorophyll.

mutated. Despite comparable expression of the CKI1(H405Q) and CKI1(D1050N) mutants with that of wild-type CKI1, neither mutant could activate the expression of *ARR6-LUC* (Fig. 1d). Furthermore, the CKI1(H405Q) mutant could exert dominant negative effects and diminish the activation of the *ARR6* promoter by exogenous *t*-zeatin. The CKI1(1050N) mutant was less potent. The results suggest that histidine protein kinase activity and phosphoryl transfer from His 405 to Asp 1050 are required for the CKI1 function in activating cytokinin signalling. The dominant negative CKI1(H405Q) mutant might interfere with cytokinin perception and/or disturb downstream signalling.

Another *Arabidopsis* hybrid histidine protein kinase, CRE1/AHK4/WOL, has been shown to be a cytokinin receptor. The evidence is based on the inability of the *cre1* mutant to respond to cytokinin in the shoot induction assay, and the ability of CRE1 to complement histidine protein kinase mutants of budding and fission yeast and *Escherichia coli* in a cytokinin-dependent manner^{6,21,23}. As CRE1/AHK4/WOL is predominantly expressed in roots^{23,24} and the *wol(cre1)* recessive mutants lack an obvious leaf phenotype^{6,23,24}, it is possible that other closely related *Arabidopsis* histidine protein kinase genes, such as *AHK2*, *AHK3* and *CKI1* (refs 3, 6, 24), provide cytokinin receptor functions. To test directly the function of these histidine protein kinases in cytokinin signalling, a construct expressing AHK2, AHK3 or CRE1(AHK4/WOL) was cotransfected into *Arabidopsis* protoplast with the cytokinin-inducible reporter *ARR6-LUC*. In contrast to CKI1, neither AHK2, AHK3 nor CRE1(AHK4/WOL) caused the activation of *ARR6-LUC* in the absence of exogenous cytokinin. However, cytokinin treatment resulted in further activation of the reporter in the presence of AHK proteins, especially CRE1(AHK4/WOL) (Fig. 1e). We then tested the requirement of the conserved His 459 and Asp 973 residues in the histidine protein kinase and receiver domains, respectively, for the CRE1 activity. CRE1(H459Q) and CRE1(D973N) mutants lost their ability to further enhance *ARR6-LUC* expression in the presence of exogenous cytokinin (Fig. 1e). Both mutants also imposed a dominant negative effect and reduced cytokinin signalling. Similar to the CKI1 mutants, the His mutant is

more potent than the Asp mutant, perhaps owing to the bimolecular phosphoryl transfer process between mutant and wild-type histidine protein kinases. These results indicate that cytokinin signals are sensed by multiple histidine protein kinase receptors with different or overlapping expression patterns²⁴. Although CRE1 is predominantly expressed in roots^{23,24}, it can certainly function in leaf cells as a cytokinin receptor. However, CKI1 and CRE1 represent two different types of cytokinin receptors that require histidine protein kinase activity and phosphoryl transfer to initiate cytokinin signalling, but have different sensing mechanisms.

To gain an insight into where cytokinin signalling is initiated, the subcellular localization of CKI1 was examined using the CKI1-green fluorescent protein (GFP) fusion. We first confirmed that CKI1-GFP acted in a similar way as CKI1-haemagglutinin (HA) in the protoplast assay based on *ARR6-LUC* activation (Fig. 1d). Confocal microscopy revealed that CKI1 is mainly localized to the plasma membrane but not in the cytosol, endoplasmic reticulum or nucleus, indicated by various GFP markers (Fig. 1f). As the amino terminus of CKI1 is loaded with putative glycosylation sites, we examined the effect of a glycosylation inhibitor, tunicamycin²⁵, in CKI1-GFP expression, localization or function. Although tunicamycin did not affect the expression of other GFP constructs (not shown), the expression of CKI1-GFP was completely abolished (Fig. 1f). This result suggests that the glycosylation of CKI1 is probably important for its stability, processing and/or trafficking in plant cells. Whether the N-terminal domain of CKI1 is important for cytokinin binding, proper folding, and/or receptor dimerization awaits further analysis.

AHP proteins as cytoplasm and nucleus shuttles

The analysis of the *Arabidopsis* genome has revealed at least five genes encoding putative AHP proteins with unknown physiological functions^{26,27}. To determine whether AHP proteins are involved in cytokinin signal transduction, AHP1, AHP2 and AHP5 were cloned into the plant expression vector and tested in the protoplast cytokinin response assay. AHP1, AHP2 and AHP5 had little effect on the expression of *ARR6-LUC* with or without *t*-zeatin (Fig. 2a).

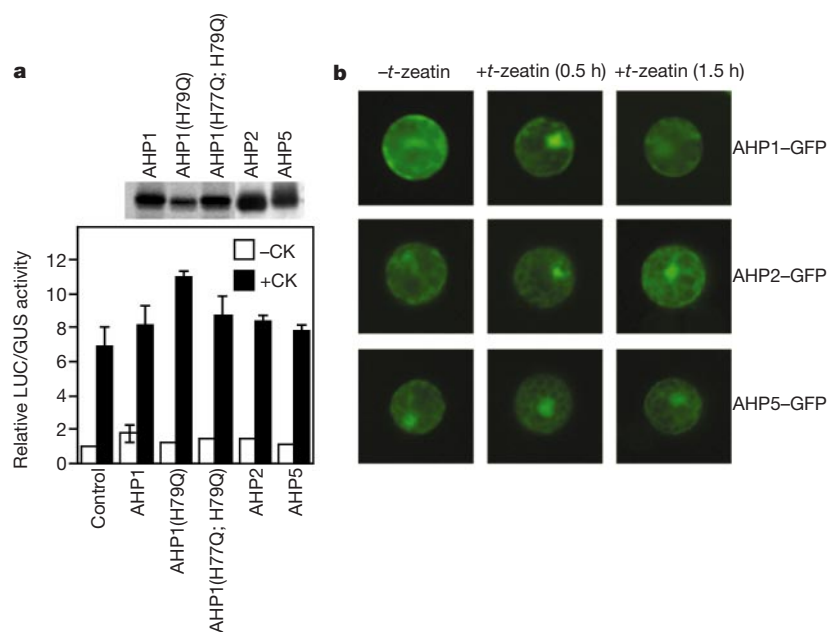


Figure 2 AHP acts as a shuttle between the cytoplasm and nucleus in cytokinin signalling. **a**, AHP overexpression does not affect cytokinin signalling. Protoplasts were cotransfected with the *ARR6-LUC* reporter and an effector plasmid expressing AHP1, AHP2, AHP5 or AHP1 mutant protein. Transfected protoplasts were treated without (–CK) or with (+CK)

100 nM *t*-zeatin. The top panel shows the expression of the AHP proteins. **b**, Cytokinin induces AHP translocation. Protoplasts were transfected with *AHP1-GFP*, *AHP2-GFP* or *AHP5-GFP* plasmid DNA. Transfected protoplasts were observed before and after *t*-zeatin (100 nM) treatment for 30–90 min.

The AHP1 mutant—where the conserved His 79 was replaced with Gln—did not repress the cytokinin activation of *ARR6-LUC*. Even the AHP1(H77Q; H79Q) double mutant was unable to exert dominant negative effect and inhibit the *ARR6-LUC* activation by cytokinin. AHP proteins and AHP1 mutants were expressed at comparable levels in the protoplast system (Fig. 2a). This result indicates that AHP proteins are not the limiting factor, or their action as potential mediators may require cytokinin-dependent modification. To further understand the role of AHP proteins in cytokinin signalling, we followed the action of AHP1-GFP in transfected protoplasts in the absence or presence of cytokinin using fluorescence microscopy. AHP1-GFP was mainly localized in the cytoplasm without cytokinin stimulation, but was translocated into the nucleus after treatment of the protoplast with *t*-zeatin. The cytokinin-dependent translocation of AHP1-GFP was transient, occurring within 30 min after cytokinin treatment (Fig. 2b). AHP2-GFP showed similar cytokinin-dependent translocation, but not AHP5-GFP, suggesting that they have different functions in plant cells.

Opposite functions of two types of nuclear ARR proteins

The requirement of histidine protein kinase activity in the initiation of cytokinin signalling from CKI1 and CRE1, the cytokinin-inducible expression of *ARR6*, the cytokinin-dependent translocation of AHP proteins into the nucleus, and the physical interaction of AHP and ARR proteins in yeast^{28–31} strongly suggest that some ARR proteins may act downstream in cytokinin signal transduction. There are two principal subfamilies of ARR genes in the *Arabidopsis* genome: the cytokinin-inducible A type and the DNA-binding B type^{11,12,31,32}. Their physiological functions are currently unclear. To investigate whether any *Arabidopsis* ARR proteins are involved in cytokinin signalling, we cloned four representative A-type ARR genes and three B-type ARR genes into the plant expression vector fused to a HA tag or GFP. As shown in Fig. 3a, A-type ARR proteins, such as ARR4, ARR5, ARR6 and ARR7, repressed *ARR6-LUC*

activity induced by 100 nM *t*-zeatin. Although ARR6 seemed to be expressed at lower abundance, its repression activity was the strongest. In contrast, the B-type ARR proteins, such as ARR1, ARR2 and ARR10, markedly activated *ARR6-LUC* expression. Ectopic expression of ARR1, ARR2 and ARR10 was sufficient to activate cytokinin signalling at different levels in the absence of exogenous cytokinin. ARR1 and ARR2 activated *ARR6-LUC* about 40- and 400-fold, respectively. Cytokinin treatment further enhanced the effect of ARR2 on *ARR6-LUC* to over 1,000-fold. The lower activation by ARR1 could be due to its lower expression level in transfected protoplasts. In the absence of cytokinin, ARR10 activated *ARR6-LUC* about 10-fold. Cytokinin enhanced the effect of ARR10 on the *ARR6-LUC* activity another 10–20-fold.

The differential effect of ARR2 and ARR10 on activating cytokinin signalling could be attributed to differences in their intrinsic activities in DNA binding and/or transcription activation. It could also be due to their distinct affinity to endogenous repressors, which probably exist in leaf protoplasts to prevent cytokinin signalling under unstimulated condition. Surprisingly, mutations of the conserved Asp residue in the receiver domains of ARR4(D95N), ARR6(D86N) and ARR2(D80N) did not alter their abilities to reduce or enhance *ARR6-LUC* expression. (Fig. 3a). This result indicates that phosphorelay stimulated by cytokinin may be involved in liberating the positive regulators such as ARR1, ARR2 and ARR10. Ectopic expression of these ARR proteins probably bypasses the negative regulation and causes constitutive cytokinin responses without the signal.

To further test the role of ARR2 in cytokinin signalling, we designed a dominant negative mutant of ARR2 (Δ ARR2), which contained only the DNA-binding domain but not the putative transcription activation and the receiver domains³². If Δ ARR2 can compete with the endogenous ARR proteins in binding to the *ARR6* promoter without transcription activation, it may block cytokinin- or CKI1-dependent activation of *ARR6-LUC*. As shown in Fig. 3b, the dominant negative mutant of ARR2 reduced *ARR6-LUC*

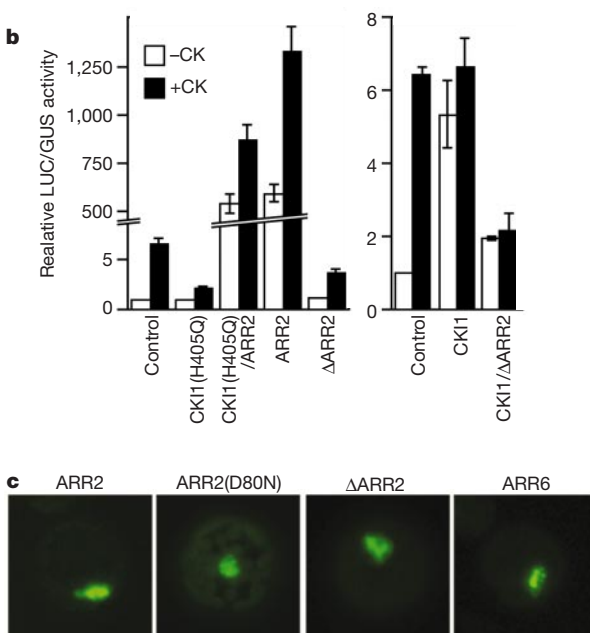
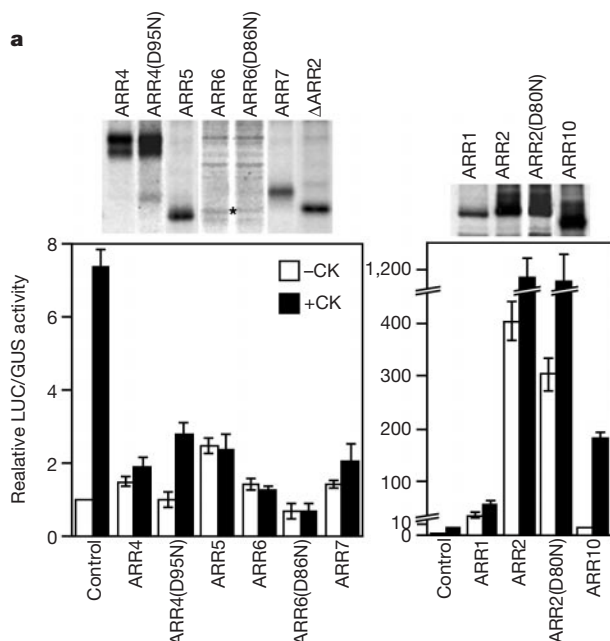


Figure 3 Opposite regulations of cytokinin primary response gene transcription by two types of ARR protein. **a**, Negative and positive regulation by ARR proteins. Protoplasts were cotransfected with the *ARR6-LUC* reporter and an effector plasmid expressing ARR1, 2, 4, 5, 6, 7, 10 or various mutant proteins, ARR2(D80N), ARR4(D95N) or ARR6(D86N). The transfected protoplasts were treated without (–CK) or with (+CK) 100 nM *t*-zeatin. The top panel shows the expression of ARR proteins. The expression of

ARR6 and ARR6(D86N) was lower than the other proteins and required longer exposure. **b**, ARR proteins act downstream of CKI1 in cytokinin signalling. Protoplasts were cotransfected with the *ARR6-LUC* reporter alone or with an effector plasmid as indicated. Δ ARR2 is a dominant negative version of ARR2. **c**, ARR proteins are localized in the nucleus. Protoplasts were transfected with *ARR2-GFP*, *ARR2(D80N)-GFP*, Δ ARR2-GFP or *ARR6-GFP* plasmid DNA and observed with a fluorescence microscope.

expression that was elicited by 100 nM *t*-zeatin or CKI1. Furthermore, the dominant negative effect of the CKI1(H405Q) mutant on the protoplast response to exogenous cytokinin could be bypassed by ectopic expression of wild-type ARR2 (Fig. 3b). This epistatic relationship places ARR2 and/or ARR2-like proteins downstream of multiple histidine protein kinase receptors in the cytokinin signal transduction pathway. The finding of multiple ARR2 binding motifs, (G/A)GAT(T/C), in the promoter regions of *ARR6* and other cytokinin-inducible genes³² suggests that ARR2 could be a master regulator in cytokinin signalling. To test further the idea that the A-type and B-type ARR proteins are transcription repressors and activators, respectively, in cytokinin signalling, we examined the subcellular localization of ARR6–GFP and ARR2–GFP. Both ARR proteins are exclusively localized in the nucleus regardless of the cytokinin treatment (Fig. 3c), consistent with nuclear localization of ARR1 and ARR2 in onion epidermal cells and parsley protoplasts^{31,32}. Their nuclear localization is probably independent of the phosphorylation state as ARR2(D80N)–GFP was found in the nucleus. The ARR2 dominant negative mutant (Δ ARR2) was also localized in the nucleus (Fig. 3c). These results indicate that the cytokinin-dependent phosphorelay does not have a role in the nuclear localization and DNA-binding of ARR proteins or their intrinsic transcription activation or repression activities.

Ectopic expression of ARR2 mimics cytokinin responses

In tissue culture, induction of cell proliferation and subsequent shoot formation require cytokinin. To determine whether the same cytokinin signalling pathway is responsible for transcription regulation, cell proliferation, shoot meristem initiation, and leaf formation, we developed a seedling cytokinin response assay. *Arabidopsis* seedlings were stably transformed using *Agrobacterium* carrying *GFP*, *CKI1*, *CKI1(H405Q)*, *ARR2* and *ARR6* constructs in the mini-binary vector pCB302 (ref. 33). Transformation with the pCB301 vector that lacks a phosphinothricin acetyl transferase (*bar*) gene did not result in viable tissues on the selection medium. The transformed seedlings with the GFP control proliferated on the indole-3-acetic acid (IAA) medium. The addition of cytokinin enhanced both cell proliferation and ectopic shoot formation; however, both CKI1 and ARR2 promoted extensive cell proliferation and shoot and leaf formation in the absence of exogenous cytokinin (Fig. 4a). Compared with the GFP control, the CKI1(H405Q) mutant and ARR6 slightly inhibited cell proliferation on the same medium with or without exogenous cytokinin (Fig. 4a). The results of this seedling cytokinin assay are consistent with those from protoplast transient expression analyses based on transcription. As the CKI1(H405Q) mutant did not promote shoot initiation and leaf formation, it is clear that the histidine protein kinase activity of CKI1 is required to initiate cytokinin signalling.

To observe the consequences of constitutive cytokinin signalling *in planta*, the binary vectors expressing CKI1, CKI1(H405Q), ARR2 and ARR6 were introduced into *Arabidopsis* plants. Ectopic expression of ARR2 in transgenic *Arabidopsis* plants delayed leaf senescence (Fig. 4b). Transgenic *Arabidopsis* plants carrying the CKI1 construct showed similar phenotype, but not the CKI1(H405Q) and ARR6 transgenic plants, which served as negative controls. Thus, CKI1 and ARR2 are positive regulators in the cytokinin signal transduction pathway that mediate transcription, cell proliferation, shoot initiation and leaf formation, as well as delaying leaf senescence. ARR2 could be one of the direct downstream targets of CKI1 and CRE1, and serve as a central regulator in diverse cytokinin responses.

Discussion

Although the conserved signature motifs of histidine protein kinases have been found in the photoreceptor phytochromes, a putative osmosensor, and ethylene and cytokinin receptors in flowering plants, the importance of histidine protein kinase activity

and phosphorelay in signal transduction remains unclear in plant cells^{6,11,13,15}. Here, we provide evidence that both histidine protein kinase activity and phosphorelay are essential for cytokinin-mediated transcription, cell proliferation, leaf formation and leaf longevity, on the basis of protoplast transient expression and transgenic tissue and plant assays. Quantitative transcription analyses (Fig. 1d, e) suggest that CKI1 and CRE1 histidine protein kinase receptors act through different cytokinin perception mechanisms. The action of CKI1 without exogenous cytokinin implies that CKI1 could be constitutively active and/or could sense endogenous signals (Fig. 5). CRE1 and perhaps AHK2 and

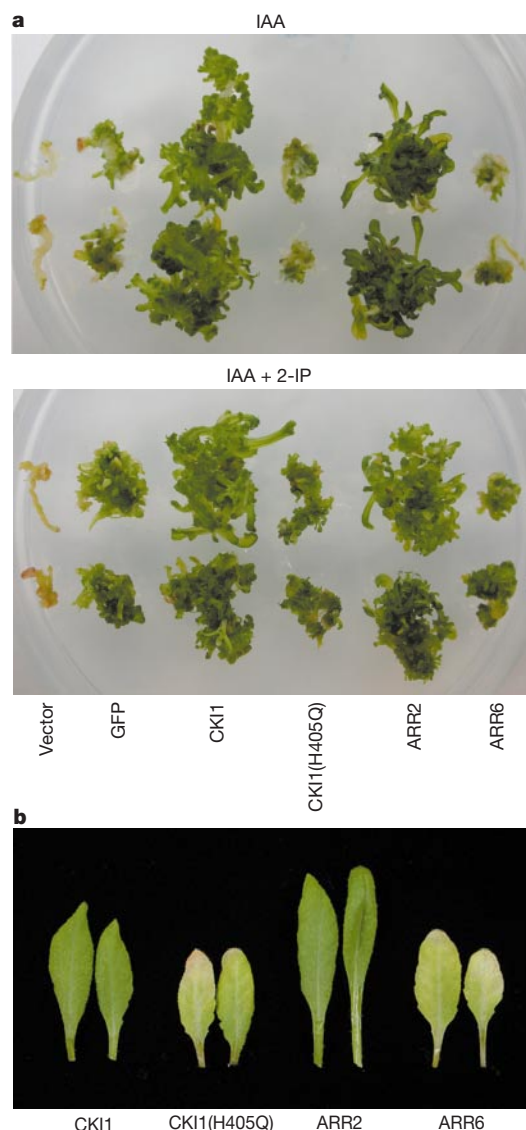


Figure 4 Ectopic expression of ARR2 is sufficient to promote cytokinin responses in transgenic tissues and plants. **a**, ARR2 stimulates cell proliferation and shoot formation in the absence of cytokinin. *Arabidopsis* seedlings were transformed with *Agrobacterium* carrying a binary vector with various constructs, including *CKI1*, *CKI1(H405Q)*, *ARR2*, *ARR6* and *GFP*. Seedlings that were transformed with the *Agrobacterium* carrying the binary vector pCB301 without a *bar* selection marker were used as a negative control. The transformed seedlings were maintained on the selection medium with auxin (IAA) only (top) or with auxin and cytokinin (2-IP) (bottom) for 14 days. The GFP control showed green callus formation with exogenous IAA but promoted shoot formation with exogenous cytokinin. **b**, Analysis of dark-induced senescence in transgenic *Arabidopsis* plants. The fully expanded fourth leaves from the transgenic plants expressing CKI1, CKI1(H405Q), ARR2 or ARR6 were detached and floated on distilled water in the dark for 4 days.

AHK3 require extracellular cytokinin for their activation (Fig. 5). Further analyses of cytokinin binding and chimaeric histidine protein kinases with swapped domains should clarify the underlying mechanism of each histidine protein kinase action in cytokinin signalling. Detailed expression analysis of each histidine protein kinase at the cellular level and the systematic isolation of knockout mutants³⁴ will help define their precise roles *in planta*.

As histidine protein kinase activity and phosphoryl transfer are required for CKI1 and CRE1 action, these histidine protein kinases probably converge on AHP proteins that serve as shuttles and phosphorelay carriers between the cytokinin receptors and the downstream nuclear responses (Fig. 5). Although other mechanisms have not been ruled out, the involvement of AHP proteins in cytokinin signalling is supported by previous studies of CKI1 interaction with AHP proteins *in vitro*^{29,35} and AHP interactions with ARR proteins in the yeast two-hybrid assay^{27–29,31}. Our analysis of AHP–GFP fusions provides the first visual and *in vivo* evidence that AHP1 and AHP2, but not AHP5, are translocated into the nucleus in a cytokinin-dependent manner (Fig. 2b). The action of AHP proteins seems to be passive and not rate-limiting because overexpression of AHP proteins does not affect cytokinin signalling (Fig. 2a). With the exception of CKI1, our studies show that simple overexpression of AHK and AHP proteins cannot markedly disturb cytokinin responses. The potential functional redundancy of AHK, AHP and ARR proteins may explain the difficulty in isolating related cytokinin response mutants based on classical genetic screen. Combining multiple knockout mutants may be necessary to reveal further cytokinin functions *in planta*.

By analogy with the SLN1 and HOG1 osmosensing pathway in yeast, it has been proposed that the ethylene receptor and osmosensor histidine protein kinases transmit signals through a MAPK cascade in *Arabidopsis*^{14,19}. The completion of the *Arabidopsis*

genome sequence has revealed a large number of AHP and ARR genes that have potential signalling roles in physiological responses of plant cells^{11,12,26–28}. Our studies unravel a cytokinin signalling circuit mediated by distinct functions of AHP and ARR proteins. The cytokinin phosphorelay mechanism is unique in that AHP proteins act as cytoplasm and nuclear shuttles, and in that enabling ARR proteins with DNA-binding domains is the nuclear consequence (Fig. 5). As the expression of ARR1, ARR2 and ARR10 is not regulated by cytokinin⁴, we propose the presence of putative repressors in controlling cytokinin responses (Fig. 5) based on the fact that ectopic expression of ARR1, ARR2 and ARR10 confers constitutive cytokinin signalling (Figs 3 and 4). The transcriptional activation of genes that encode the repressor-type of ARR proteins (Fig. 3a) possibly provides a negative feedback loop in controlling the transient induction of cytokinin primary response genes, and allows resetting and/or fine-tuning of the physiological state of the cells.

The cellular assay based on cytokinin-inducible transcription has identified the transcription activators ARR1, ARR2 and ARR10 as the central rate-limiting step in cytokinin signalling. These factors could potentially be regulated by signals other than cytokinin³⁶ to provide a crosstalk mechanism in plant signalling networks. ARR2 seems to act as a master regulator, manifested by its ability to mimic a broad spectrum of cytokinin actions in transgenic *Arabidopsis*. Other B-type ARRs could share a similar role. Further experiments will be required to elucidate the details in cytokinin perception and protein–protein interactions that are essential in cytokinin signalling. The expression analysis of an ARR5–GUS transgene in *Arabidopsis* has shown that cytokinin responses can occur in many different cell types, including leaf cells⁵. The short cytokinin signalling circuit, elucidated by using the mesophyll protoplast transient expression assay, could represent a conserved core signalling pathway in different cell types in response to cytokinin. This hypothesis is supported by the similar effects of CKI1, ARR2 and their mutants on cytokinin signalling analysed in mesophyll protoplasts and in stably transformed tissues and plants. However, additional cell-type-specific components probably have important roles for cytokinin responses in different cell types and tissues, for example dividing and non-dividing cells. The expression patterns of various histidine protein kinases, AHP and ARR proteins may also contribute to their unique or redundant roles in cytokinin responses^{24,26,31,32,37}. It will be of interest to investigate how this cytokinin signal transduction pathway influences cell cycle^{5,38,39}, leaf senescence^{40–42}, and shoot initiation and leaf patterning controlled by transcription factors⁴³. The combination of the complete *Arabidopsis* genome sequence and the development of the physiological leaf cell assays will provide useful tools to elucidate other histidine protein kinase-mediated signal transduction pathways in plants. □

Methods

Plasmid constructs

The 2.4-kb *Arabidopsis* ARR6 promoter was amplified by polymerase chain reaction (PCR) and fused to the firefly luciferase gene to create the ARR6–LUC reporter construct. We amplified CKI1 and AHK2 genes from the *Arabidopsis* genomic DNA by PCR. The AHK3, CRE1 (CRE1/WOL), AHP1, 2, 5 and ARR1, 2, 4, 5, 6, 7 and 10 coding regions were obtained by PCR from an *Arabidopsis* complementary DNA library⁴⁴. All of the mutants were generated by QuickChange site-directed mutagenesis (Stratagene). The coding regions of all proteins were tagged with either two copies of the haemagglutinin epitope (DHA) or GFP and inserted into a plant expression vector that contained the 35S4PPDK promoter⁴⁴ and the NOS terminator. We confirmed all PCR products and the mutations by DNA sequencing.

Arabidopsis protoplast transient expression assay

Arabidopsis Bensheim protoplasts were isolated and transfected as described with some modifications⁴⁵. Typically, 2×10^4 protoplasts were transfected with 20 µg plasmid DNA with different combinations of reporter, effector and internal control. Transfected protoplasts were incubated at 1×10^4 per ml with or without 1–100 nM *t*-zeatin for 3–6 h under light conditions at 23 °C. The UBQ10–GUS construct was used as an internal

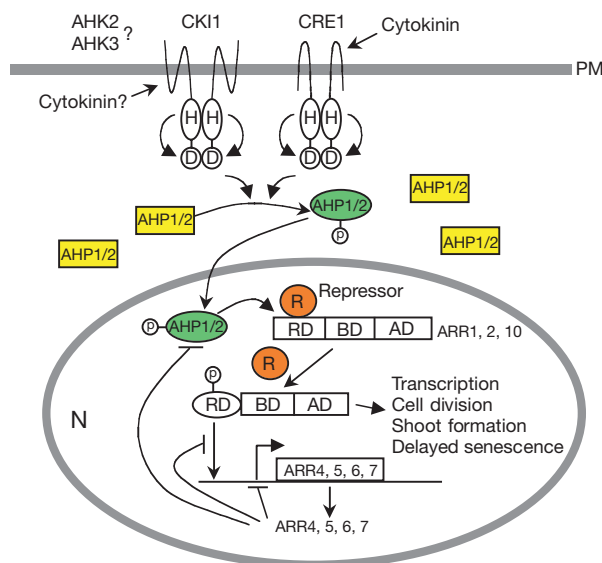


Figure 5 Model of the cytokinin signal transduction pathway in *Arabidopsis*. Cytokinin signal is perceived externally or internally by multiple histidine protein kinases at the plasma membrane. On perception of the cytokinin signal, histidine protein kinases initiate a signalling cascade through the phosphorelay that results in the nuclear translocation of AHP proteins from the cytosol. Activated AHP proteins interact with sequestered ARR proteins or ARR complexes, and release the activation type of ARR proteins from putative repressors in the nucleus. The liberated ARR proteins bind to multiple *cis* elements in the promoter of target genes. The transcription activation is essential for cell proliferation, shoot formation and delayed leaf senescence. Activation of the repressor-type of ARR genes as cytokinin primary response genes provides a negative feedback mechanism. RD, response domain; BD, DNA-binding domain; AD, transcription activation domain; PM, plasma membrane; N, nucleus; R, putative repressor; H, histidine; D, aspartate.

control to normalize the variations of each transfection in cell numbers, transformation efficiency and cell viability. The results are shown as the means of relative LUC activities from duplicate samples with error bars. All transient experiments were repeated at least three times with similar results. GFP fluorescence was observed by either Nikon TE200 fluorescent microscopy or Leica TCSNT confocal microscopy.

Protein immunoprecipitation

Transfected protoplasts were incubated with ^{35}S -labelled methionine ($200\ \mu\text{Ci ml}^{-1}$) for 6 h. The effector proteins were immunoprecipitated as described⁴⁴, analysed by SDS–polyacrylamide gel electrophoresis (PAGE) (10 or 12.5%) and visualized by fluorography.

Seedling cytokinin response assay

Four-day-old *Arabidopsis* Bensheim seedlings were placed on the callus induction medium, containing $0.5\ \text{mg l}^{-1}$ 2,4-dichlorophenoxy acetic acid, $0.05\ \text{mg l}^{-1}$ BA, and $0.05\ \text{mg l}^{-1}$ kinetin, for 4 days. The seedlings were co-cultivated and transformed with *Agrobacterium* GV3101 carrying pCB302 with either *CK11*, *CK11(H405Q)*, *ARR2*, *ARR6*, *GFP* under control of the *35SCAPDK* promoter, or the pCB301 binary vector³³ alone for three days. Transformed seedlings were then selected on the glufosinate ammonium-containing medium (with either $0.15\ \text{mg l}^{-1}$ IAA alone or with $0.15\ \text{mg l}^{-1}$ IAA and $5\ \text{mg l}^{-1}$ 2-IP) and observed up to 4 weeks. The transformed proliferating tissues were mostly derived from the apical shoot meristem of the seedlings.

Transgenic plant analysis

The same constructs that were analysed in the *Arabidopsis* protoplast transient expression and seedling assays, including *CK11*, *CK11(H405Q)*, *ARR2* and *ARR6*, were used to generate *Arabidopsis* transgenic plants using the floral dip method and bar selection as described⁴⁶. We obtained more than 200 transgenic plants with each construct. For senescence assay, fully expanded fourth leaves from representative transgenic plants were detached and floated on distilled water in the dark for 4 days.

Received 11 June; accepted 8 August 2001.

1. Davies, P. J. *Plant Hormones: Physiology, Biochemistry and Molecular Biology* (ed. Davies, P. J.) (Kluwer Academic, The Netherlands, 1995).
2. Mok, D. W. & Mok, M. C. Cytokinin metabolism and action. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 89–118 (2001).
3. Kakimoto, T. CK11, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* **274**, 982–985 (1996).
4. Kiba, T. *et al.* Differential expression of genes for response regulators in response to cytokinins and nitrate in *Arabidopsis thaliana*. *Plant Cell Physiol.* **40**, 767–771 (1999).
5. D'Agostino, I. B., Deruere, J. & Kieber, J. J. Characterization of the response of the *Arabidopsis* senescence regulator gene family to cytokinin. *Plant Physiol.* **124**, 1706–1717 (2000).
6. Inoue, T. *et al.* Identification of CRE1 as a cytokinin receptor from *Arabidopsis*. *Nature* **409**, 1060–1063 (2001).
7. Sakakibara, H., Taniguchi, M. & Sugiyama, T. His-Asp phosphorelay signaling: a communication avenue between plants and their environment. *Plant Mol. Biol.* **42**, 273–278 (2000).
8. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
9. Thomason, P. & Kay, R. Eukaryotic signal transduction via histidine-aspartate phosphorelay. *J. Cell Sci.* **113**, 3141–3150 (2000).
10. Wurgler-Murphy, S. M. & Saito, H. Two-component signal transducers and MAPK cascades. *Trends Biochem. Sci.* **22**, 172–176 (1997).
11. Urao, T., Yamaguchi-Shinozaki, K. & Shinozaki, K. Two-component systems in plant signal transduction. *Trends Plant Sci.* **5**, 67–74 (2000).
12. D'Agostino, I. B. & Kieber, J. J. Phosphorelay signal transduction: the emerging family of plant response regulators. *Trends Biochem. Sci.* **24**, 452–456 (1999).
13. Yeh, K. C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
14. Urao, T. *et al.* A transmembrane hybrid-type histidine kinase in *Arabidopsis* functions as an osmosensor. *Plant Cell* **11**, 1743–1754 (1999).
15. Bleeker, A. B. & Kende, H. Ethylene: a gaseous signal molecule in plants. *Annu. Rev. Cell Dev. Biol.* **16**, 1–18 (2000).
16. Maeda, T., Wurgler-Murphy, S. M. & Saito, H. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242–245 (1994).
17. Posas, F. *et al.* Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 'two-component' osmosensor. *Cell* **86**, 865–875 (1996).
18. Posas, F. & Saito, H. Activation of the yeast SSK2 MAP kinase kinase by the SSK1 two-component response regulator. *EMBO J.* **17**, 1385–1394 (1998).
19. Clark, K. L., Larsen, P. B., Wang, X. & Chang, C. Association of the *Arabidopsis* CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors. *Proc. Natl Acad. Sci. USA* **95**, 5401–5406 (1998).
20. Gamble, R. L., Coonfield, M. L. & Schaller, G. E. Histidine kinase activity of the ETR1 ethylene receptor from *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95**, 7825–7829 (1998).

21. Suzuki, T. *et al.* The *Arabidopsis* sensor his-kinase, *ahk4*, can respond to cytokinins. *Plant Cell Physiol.* **42**, 107–113 (2001).
22. Frank, M. *et al.* Hormone autotrophic growth and differentiation identifies mutant lines of *Arabidopsis* with altered cytokinin and auxin content or signaling. *Plant Physiol.* **122**, 721–729 (2000).
23. Mahonen, A. P. *et al.* A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev.* **14**, 2938–2943 (2000).
24. Ueguchi, C., Koizumi, H., Suzuki, T. & Mizuno, T. Novel family of sensor histidine kinase genes in *Arabidopsis thaliana*. *Plant Cell Physiol.* **42**, 231–235 (2001).
25. Elbein, A. D. Inhibitors of the biosynthesis and processing of N-linked oligosaccharides. *CRC Crit. Rev. Biochem.* **16**, 21–49 (1984).
26. Suzuki, T., Imamura, A., Ueguchi, C. & Mizuno, T. Histidine-containing phosphotransfer (Hpt) signal transducers implicated in His-to-Asp phosphorelay in *Arabidopsis*. *Plant Cell Physiol.* **39**, 1258–1268 (1998).
27. Suzuki, T. *et al.* Compilation and characterization of histidine-containing phosphotransmitters implicated in His-to-Asp phosphorelay in plants: AHP signal transducers of *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* **64**, 2486–2489 (2000).
28. Imamura, A. *et al.* Compilation and characterization of *Arabidopsis thaliana* response regulators implicated in His-Asp phosphorelay signal transduction. *Plant Cell Physiol.* **40**, 733–742 (1999).
29. Urao, T., Miyata, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. Possible His to Asp phosphorelay signaling in an *Arabidopsis* two-component system. *FEBS Lett.* **478**, 227–232 (2000).
30. Suzuki, T., Sakurai, K., Ueguchi, C. & Mizuno, T. Two types of putative nuclear factors that physically interact with histidine-containing phosphotransfer (Hpt) domains, signaling mediators in His-to-Asp phosphorelay, in *Arabidopsis thaliana*. *Plant Cell Physiol.* **42**, 37–45 (2001).
31. Lohrmann, J. *et al.* The response regulator ARR2: a pollen-specific transcription factor involved in the expression of nuclear genes for components of mitochondrial complex I in *Arabidopsis*. *Mol. Genet. Genomics* **265**, 2–13 (2001).
32. Sakai, H., Aoyama, T. & Oka, A. *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. *Plant J.* **24**, 703–711 (2000).
33. Xiang, C., Han, P., Lutziger, L., Wang, K. & Oliver, D. J. A mini binary vector series for plant transformation. *Plant Mol. Biol.* **40**, 711–717 (1999).
34. Sussman, M. R., Amasino, R. M., Young, J. C., Krysan, P. J. & Austin-Phillips, S. The *Arabidopsis* knockout facility at the University of Wisconsin-Madison. *Plant Physiol.* **124**, 1465–1467 (2000).
35. Nakamura, A. *et al.* Biochemical characterization of a putative cytokinin-responsive His-kinase, CK11, from *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* **63**, 1627–1630 (1999).
36. Urao, T., Yakubov, B., Yamaguchi-Shinozaki, K. & Shinozaki, K. Stress-responsive expression of genes for two-component response regulator-like proteins in *Arabidopsis thaliana*. *FEBS Lett.* **427**, 175–178 (1998).
37. Sakai, H., Aoyama, T., Bono, H. & Oka, A. Two-component response regulators from *Arabidopsis thaliana* contain a putative DNA-binding motif. *Plant Cell Physiol.* **39**, 1232–1239 (1998).
38. Riou-Khamlichi, C., Huntley, R., Jacquemard, A. & Murray, J. A. Cytokinin activation of *Arabidopsis* cell division through a D-type cyclin. *Science* **283**, 1541–1544 (1999).
39. Meijer, M. & Murray, J. A. Cell cycle controls and the development of plant form. *Curr. Opin. Plant Biol.* **4**, 44–49 (2001).
40. Gan, S. & Amasino, R. M. Inhibition of leaf senescence by autoregulated production of cytokinin. *Science* **270**, 1986–1988 (1995).
41. Ori, N. *et al.* Leaf senescence is delayed in tobacco plants expressing the maize homeobox gene *knotted1* under the control of a senescence-activated promoter. *Plant Cell* **11**, 1073–1080 (1999).
42. Quirino, B. F., Noh, Y. S., Himelblau, E. & Amasino, R. M. Molecular aspects of leaf senescence. *Trends Plant Sci.* **5**, 278–282 (2000).
43. Tsiantis, M. Control of shoot cell fate. Beyond homeoboxes. *Plant Cell* **13**, 733–738 (2001).
44. Sheen, J. Ca^{2+} -dependent protein kinases and stress signal transduction in plants. *Science* **274**, 1900–1902 (1996).
45. Kovtun, Y., Chiu, W. L., Tena, G. & Sheen, J. Functional analysis of oxidative stress-activated mitogen-activated protein kinase cascade in plants. *Proc. Natl Acad. Sci. USA* **97**, 2940–2945 (2000).
46. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
47. Haseloff, J., Siemering, K. R., Prasher, D. C. & Hodge, S. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc. Natl Acad. Sci. USA* **94**, 2122–2127 (1997).
48. Yanagisawa, S. & Sheen, J. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* **10**, 75–89 (1998).

Acknowledgements

We would like to thank T. Kakimoto for the *CK11* cDNA clone and communicating unpublished results; C. Xiang, K. Wang and D. J. Oliver for the mini-binary vectors; J. Callis for the *UBQ10–GUS* plasmid; E. Schaller for the *ER–GFP* construct; J. Elhai, W.-L. Chiu and H. Sze for helpful discussions on the manuscript; B. Moore for help with confocal microscopy; and G. Tena for *Arabidopsis* plants. This work is supported by NSF and NIH grants to J.S.

Correspondence and requests for materials should be addressed to J.S. (e-mail: sheen@molbio.mgh.harvard.edu).

The landing of the NEAR-Shoemaker spacecraft on asteroid 433 Eros

J. Veverka*, B. Farquhar†, M. Robinson‡, P. Thomas*, S. Murchie†, A. Harch*, P. G. Antreasian§, S. R. Chesley§, J. K. Miller§, W. M. Owen Jr§, B. G. Williams§, D. Yeomans§, D. Dunham†, G. Heyler†, M. Holdridge†, R. L. Nelson†, K. E. Whittenburg†, J. C. Ray†, B. Carcich*, A. Cheng†, C. Chapman||, J. F. Bell III*, M. Bell*, B. Bussey‡, B. Clark*, D. Domingue†, M. J. Gaffey§, E. Hawkins†, N. Izenberg†, J. Joseph*, R. Kirk#, P. Lucey*, M. Malin**, L. McFadden††, W. J. Merline||, C. Peterson*, L. Prockter†, J. Warren† & D. Wellnitz††

* Space Sciences Building, Cornell University, Ithaca, New York 14853, USA

† Applied Physics Laboratory, Johns Hopkins University,

1110 Johns Hopkins Road, Laurel, Maryland 20723, USA

‡ Department of Geological Sciences, Northwestern University, 309 Locy Hall, Evanston, Illinois 60208, USA

§ Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

|| Southwest Research Institute, 1050 Walnut Street, Suite 426, Boulder, Colorado 80302, USA

¶ Department of Earth and Environmental Sciences, Science Center, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180-3590, USA

US Geological Survey, 2255 N. Gemini Drive, Flagstaff, Arizona 86001, USA

* Hawaii Institute of Geophysics and Planetology, University of Hawaii, 2525 Correa Road, Honolulu, Hawaii 96822, USA

** Malin Space Science Systems, Inc., PO Box 910148, San Diego, California 92191, USA

†† Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

The NEAR-Shoemaker spacecraft was designed to provide a comprehensive characterization of the S-type asteroid 433 Eros (refs 1–3), an irregularly shaped body with approximate dimensions of $34 \times 13 \times 13$ km. Following the completion of its year-long investigation, the mission was terminated with a controlled descent to its surface, in order to provide extremely high resolution images. Here we report the results of the descent on 12 February 2001, during which 70 images were obtained. The landing area is marked by a paucity of small craters and an abundance of 'ejecta blocks'. The properties and distribution of ejecta blocks are discussed in a companion paper⁴. The last sequence of images reveals a transition from the blocky surface to a smooth area, which we interpret as a 'pond'. Properties of the 'ponds' are discussed in a second companion paper⁵. The closest image, from an altitude of 129 m, shows the interior of a 100-m-diameter crater at 1-cm resolution.

The main goal of the Near Earth Asteroid Rendezvous (NEAR) landing was to capture images of Eros at extremely high resolution, potentially improving the resolution by a factor of 10 over the best images (resolutions of about 0.3 m per pixel) acquired during the low-altitude flyover on 28 January 2001 (ref. 6). The imaging sequence consisted of two parts. First, two 6-frame monochrome mosaics were acquired following the first of five braking manoeuvres and used by the navigation team to calculate the timing shift for the final burns and the initiation of the final imaging sequence. The second and primary part of the descent imaging consisted of a continuous set of frames that began just before the start of the second braking manoeuvre and continued uninterrupted through the three remaining burns until 15 min after the predicted time of contact with the surface, to allow for the latest probable landing time in the event of over-burns that would cause the spacecraft to stay aloft longer.

The sequence contained monochrome images (760 nm) with image spacing that alternated between 20 and 45 s. Images were returned to Earth immediately using a new method in which the

images were buffered in real-time and then played back while the next set was being acquired. All images were shuttered using an exposure of 15 ms to limit smear. Here we summarize the major insights about the surface of Eros obtained from the images. More detailed analyses of specific observations are provided in accompanying contributions describing ponded deposits⁵ and ejecta blocks⁴.

The selection of the landing site was dictated by several practical considerations that are summarized in the legend to Fig. 1. One aim was to land the NEAR spacecraft in a region where details of surface modification by ejecta and downslope movement of regolith could be studied. We hoped to land in the vicinity of one of the enigmatic 'ponds' discovered during the 28 October 2000 low-altitude flyby⁶ and viewed at even better resolution (0.5 m) in late January 2001. The location of the landing site, within the 9-km depression Himeros, is shown in Fig. 1 (arrows). Several reconstructions of the final trajectory yield estimates of the landing location, which differ by several hundred metres on the surface of Eros. The best solution, supported by the appearance of the surface in the final images, indicates that NEAR landed inside a degraded crater about 100 m in diameter at 35.7° S, 279.5° W. A strong supporting

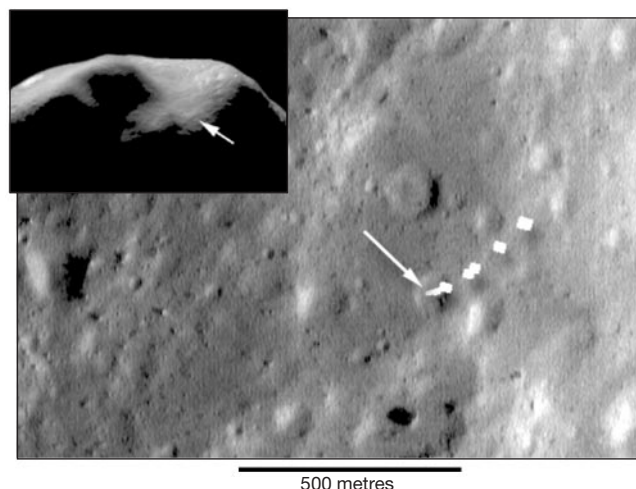


Figure 1 The NEAR landing site (35.7° S, 279.5° W). Outlines of the last ten frames are indicated. The final four images (Fig. 3) fall inside the 100-m-diameter crater indicated by the arrow. The descent trajectory was designed to maximize the number of images of the surface from altitudes below 5 km. Minimizing the impact velocity was a secondary goal. To maximize image downlink capability, the spacecraft had to maintain constant high-gain antenna contact with the Earth. The NEAR-Shoemaker (NEAR) spacecraft has fixed instruments, radio antennas and solar arrays. This limited the possibilities for imager pointing. Because the rotation pole of Eros lies almost in the plane of the asteroid's orbit about the Sun, the solar latitude during February 2001 was at 85° S, with southern latitudes in constant sunlight and northern regions in darkness. Simulations showed that descent trajectories to landing sites along the smaller axis of Eros were less sensitive to orbit determination timing errors than to those on the long axis. The longitude of the touchdown site was selected so that the spacecraft could maintain continuous Earth contact and have the imager pointed at the surface of Eros during descent. Before the descent, the NEAR spacecraft was in a near-circular 34 km by 36 km retrograde orbit. A de-orbit burn of 2.57 m s^{-1} performed on 12 February at 15:14 UTC changed the orbit inclination from 180 degrees to 135 degrees relative to Eros' equator. Four additional braking manoeuvres were pre-programmed to execute at fixed intervals during the 4.5-h controlled descent. The time of impact from Doppler tracking was determined to be 19:44:16 UTC. Post-landing analysis indicated a vertical impact velocity of 1.5 to 1.8 m s^{-1} and a transverse impact velocity of 0.1 m s^{-1} to 0.3 m s^{-1} . The touchdown site was determined to be at 35.7° S, 279.5° W, about 500 m from the nominal site. All times are spacecraft event times in UTC (Coordinated Universal Time). On 12 February 2001, Eros was 2.11 astronomical units (AU) from Earth corresponding to a one-way light time from the spacecraft to Earth of 17 min 34.5 s. Inset, a global view of Eros showing the landing site within the large 9-km depression Himeros. The inset spans 17 km.

argument is that the last image shows that the spacecraft landed on or within a few metres of a pond, a landform known to occur predominantly on the floors of craters^{5,6}.

At the beginning of the descent-imaging sequence, the geometry was non-optimal for discerning details of surface morphology (high Sun, oblique viewing) but improved as the spacecraft approached the surface. The last views were obtained under ideal illumination (Sun 26° above the horizon) looking almost vertically (16° off vertical). In all, 70 images were obtained during the final sequence over about 37 minutes beginning at a range to the surface of about 7 km (resolution of 70 cm per pixel) and ending at 129 m

(resolution of 1 cm per pixel). The camera remained in focus throughout.

As expected from previous lower-resolution views⁶, the landing region displayed a subdued appearance, with degraded and apparently filled-in craters dominating the landscape, with abundant evidence of ejecta blocks and regolith. The nature and the size distribution of the ejecta blocks is discussed in detail in ref. 4.

As the spacecraft reached altitudes below 1 km (resolution of 10 cm per pixel or better) local differences in surface texture became evident. The morphology of the surface remained generally blocky

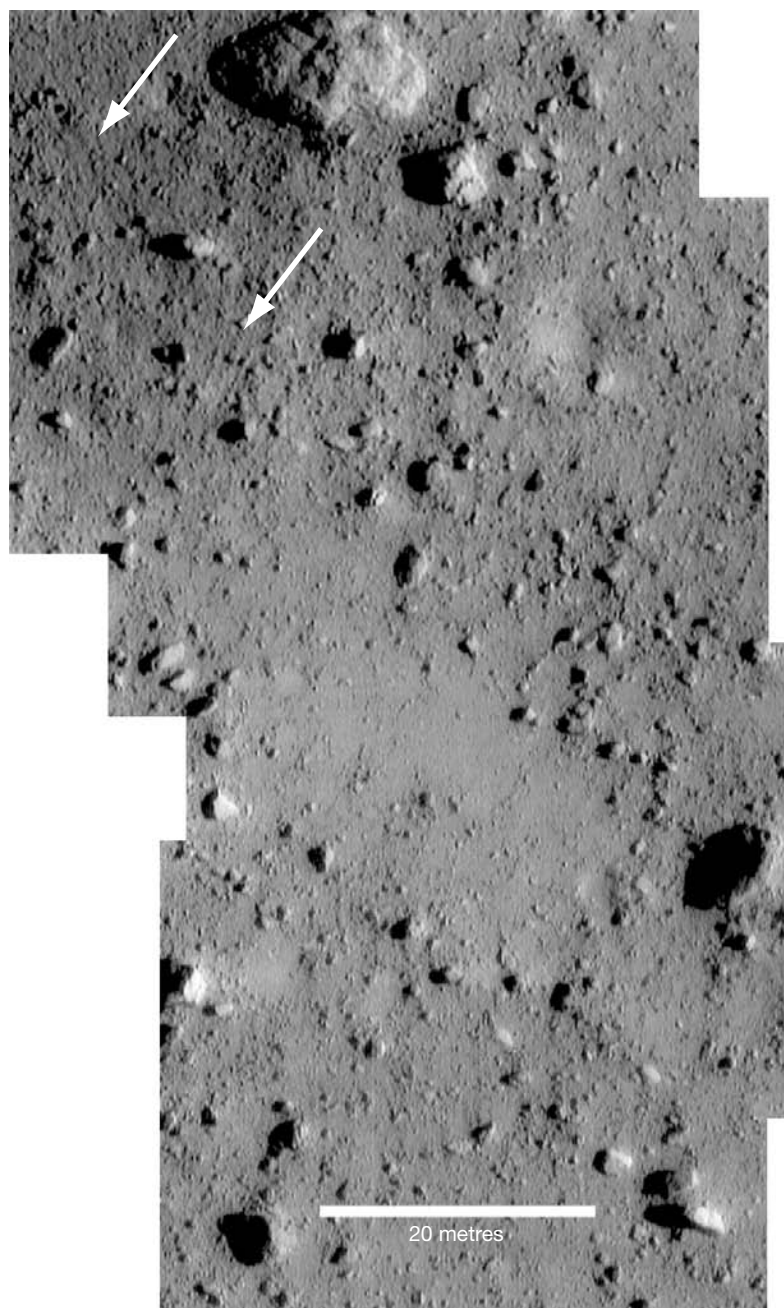


Figure 2 Details of regolith morphology at high resolution. White arrows point to furrow-like lineations with some sinuosity trending from top to bottom in the general direction of the smooth area near the middle of the mosaic. This direction coincides with the general trend of numerous apparent alignments of small rocks and positive relief features at the limit of resolution throughout the mosaic. Images MET 157416658, -678 and -743. Average range, 900 m (resolution of 90 cm per pixel). The NEAR camera,

the MSI, covers the spectral range from 400 to 1,000 nm and has a 2.25° by 2.90° field of view^{13,14}. The 244 by 537 charge-coupled device (CCD) has rectangular pixels that subtend 0.95 m by 1.61 m at 10 km. All descent images were obtained through filter 3 centred at 760 nm. Resolutions quoted in this paper correspond to the short dimension of the pixel.

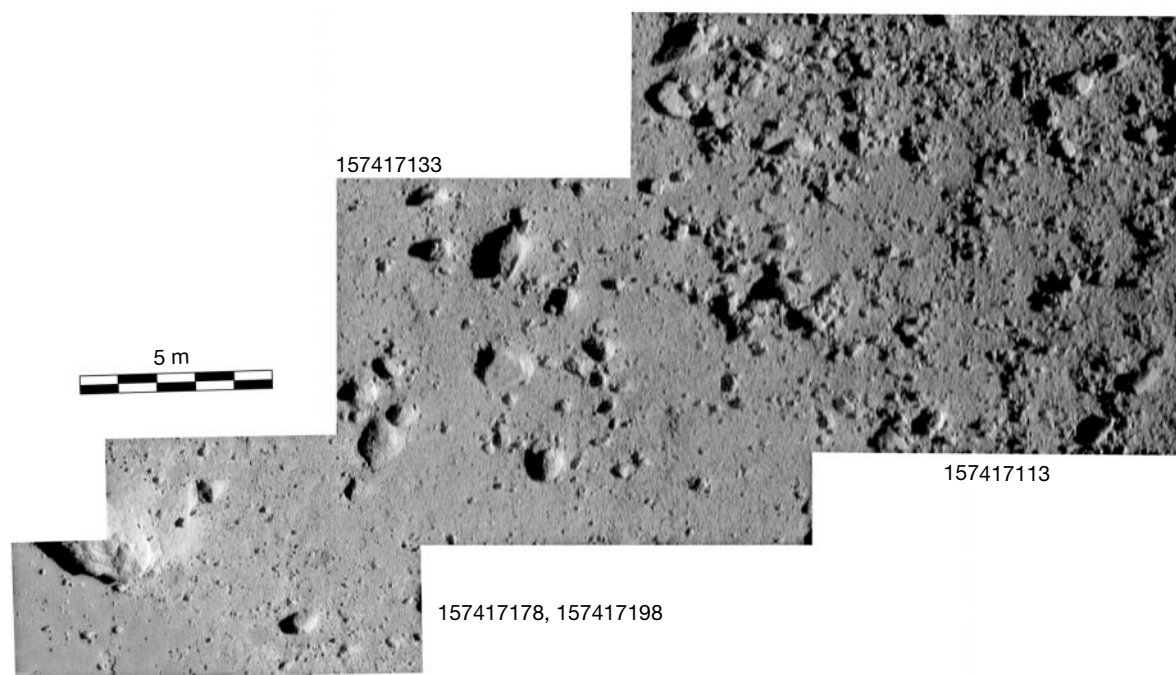


Figure 3 Last four images of the descent sequence, showing the floor of the crater indicated by the arrow in Fig. 1. A gradual transition from a blocky morphology (upper right) to a smooth 'pond' deposit (bottom left) is evident. A 5-m scale bar is included. The best estimate places the actual landing site some 7 m to the left edge of the mosaic. Mosaic of frames MET 157417113, -33, -78 and -98. The spacecraft (with solar panels deployed) is about 6 m across. Although telemetry ceased when the spacecraft landed, carrier lock was maintained indicating that NEAR had survived and was still

operational. Telemetry was later restored, and it was determined that NEAR was resting on the tips of two solar panels and the bottom edge of the spacecraft's body. Although the mission was scheduled to end on 14 February 2001, NASA decided to extend the mission by 14 days to gather additional γ -ray spectrometer and magnetometer data. On 28 February, commands were sent to place the spacecraft into a hibernational mode. NASA may try to contact NEAR one more time in September 2002 when NEAR is back in sunlight and Eros is only 0.64 AU from Earth.

but smoother, more block-free areas appeared (Fig. 2). Some of these relatively block free areas have smooth textures, while others appear to be furrowed by linear markings which have characteristics suggestive of downslope displacement of surface material. The sinuous furrows (arrows in Fig. 2) trend in a direction that would move material into the flat, smoother area in the middle of the mosaic of images. This direction coincides with the trend of apparent alignments of small rocks and other positive relief features. One problem is that views from other directions and under different lighting conditions are not available, making it difficult to exclude the remote possibility that some of the apparent alignments are artefacts of the shadowing under the prevailing specific lighting and viewing conditions⁷. Rock fragments down to scales of tens of centimetres are seen to be buried partially in a finer regolith. In many cases very intricate textures can be discerned on the surfaces of the blocks, including pitting, planar fractures and at least one example of a small crater.

The very last sequence of images (Fig. 3) shows a definite transition from a blocky surface (upper right) to a very smooth area (lower left). We interpret the smooth area as a 'pond' and consider the observed transition as a strong clue that the spacecraft landed inside a crater^{5,6}. Several unusual morphological features can be seen in the last images (Fig. 4). First, the very smooth (at a scale of 1 cm) texture of the pond deposit at lower left. Second, the occurrence of several collapse pits within the deposit (arrows a and b in Fig. 4A). The smaller ones are about 5 cm across. The largest (arrow a) has several stubby extensions. In all cases the morphology of these pits suggests collapse by removal of subsurface support. Their depth can be estimated at 1 to 2 cm.

All of the larger ejecta blocks in this region are partially buried. In a few cases there may be debris aprons around individual rocks and in one case (arrow c in Fig. 4A) there is clearly a moat-like depression around part of the periphery of a 1-m boulder. Such

depressions suggest compaction of a fine textured material, perhaps by a process related to the one responsible for the formation of the pits. There are hints of subtle channel-like depressions (arrow d in Fig. 4A) trending from upper right to lower left in a direction consistent with transporting fine material into the pond area. The transition between the very smooth pond deposit and the surrounding area is fairly abrupt, occurring on a scale of 10 to 20 cm.

It is difficult to compare the detailed regolith morphology of Eros with observations of other asteroids owing to the lack of images of a comparable resolution. The best Galileo views of S-asteroids Gaspra and Ida had resolutions of 60 m per pixel and 30 m per pixel, respectively and were limited to only small fractions of the surfaces^{8,9}. The lack of images at metre resolution of Gaspra and Ida precludes any useful comparisons of the average surface density of ejecta blocks or the size distribution of blocks on Eros and the two S-type asteroids studied by Galileo. Data for Mars' satellite Phobos at resolutions of about 3 m per pixel show block populations in the vicinity of the large crater Stickney which are comparable to those on Eros¹⁰.

Nothing comparable to the flat 'pond' deposits has been noted on Gaspra, Ida or Phobos, even though Phobos coverage is certainly adequate to show such features if they were present¹⁰. Filling in of depressions by finer regolith has been observed on Mars' satellite Deimos but the morphology of such deposits is very different from that of the ponds on Eros¹¹.

A most remarkable characteristic of the area in which NEAR landed is the scarcity of small craters. The cause remains to be determined¹², but could be related to the transport of regolith into local sedimentary basins⁵. Alternatively it could be due to blanketing by ejecta from a relatively recent large impact⁴. The preservation of copious ejecta blocks suggests that in recent epochs in this region of Eros small craters are more likely to be obliterated by burial than by erosion. □

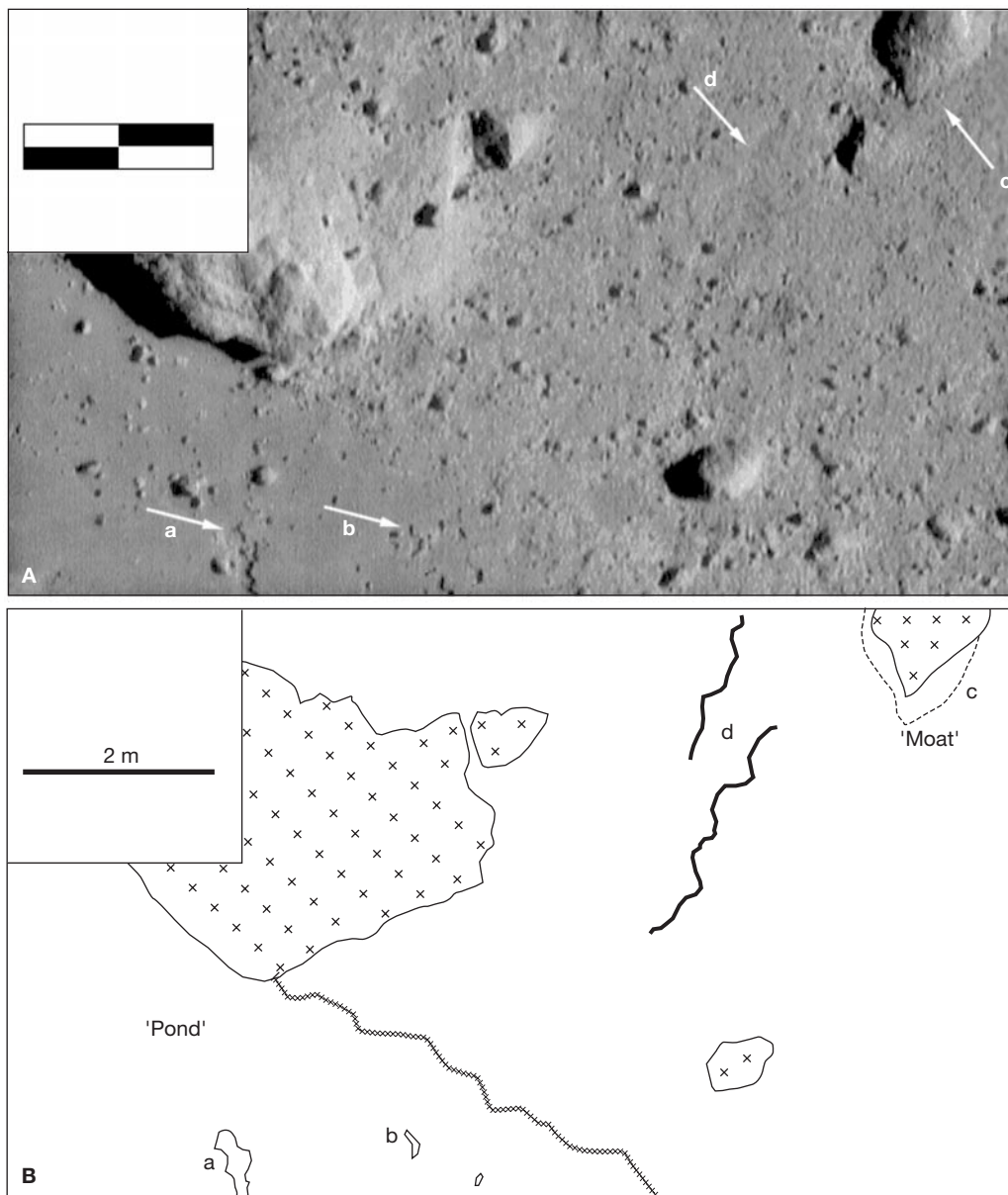


Figure 4 Geological interpretation of the landing site. **A**, NEAR's last two images: details of pond with collapse features. These frames have an average resolution of 1.4 cm. Two enclosed collapse pits are indicated by arrows (a and b) in the smooth pond deposit. Arrow c points to a moat-like depression around a metre-sized boulder. Arrow d marks a subdued, somewhat sinuous, channel-like depression which trends in the general

direction of the pond deposit. Mosaic of frames MET 157417178 and -98. Range and resolutions are 167 m (1.6 cm per pixel) and 129 m (1.2 cm per pixel), respectively. Scale bar, 2 m. **B**, Sketch map of **A**. Blocks are outlined with cross filling. Boundary between pond and rougher surface is marked by crossed line. Two sinuous depressions are shown by heavy, solid lines (d). Collapse features are outlined within the pond (a and b).

Received 18 June; accepted 6 August 2001.

- Farquhar, R. W. *et al.* NEAR mission overview and trajectory design. *J. Astronaut. Sci.* **43**, 353–372 (1999).
- Veverka, J. *et al.* NEAR at Eros: Imaging and spectral results. *Science* **289**, 2088–2097 (2000).
- Zuber, M. T. *et al.* The shape of 433 Eros from the NEAR-Shoemaker laser rangefinder. *Science* **289**, 2097–2101 (2000).
- Thomas, P. C., Veverka, J., Robinson, M. S. & Murchie, S. Shoemaker crater as the source of most ejecta blocks on the asteroid 433 Eros. *Nature* **413**, 394–396 (2001).
- Robinson, M. S., Thomas, P. C., Veverka, J., Murchie, S. & Carcich, B. The nature of ponded deposits on Eros. *Nature* **413**, 396–400 (2001).
- Veverka, J. *et al.* Imaging of small-scale features on 433 Eros from NEAR: Evidence for a complex regolith. *Science* **292**, 484–488 (2001).
- Wolfe, E. M. & Bailey, N. G. Lineaments of the Apennine Front-Apollo 15 landing site. *Proc. Lunar Planet. Sci. Conf.* **3**, 15–25 (1972).
- Sullivan, R. *et al.* Geology of 243 Ida. *Icarus* **120**, 119–139 (1996).
- Lee, P. *et al.* Ejecta blocks on 243 Ida and on other asteroids. *Icarus* **120**, 87–105 (1996).
- Thomas, P. C. *et al.* Phobos: Regolith and ejecta blocks investigated with Mars orbiter camera images. *J. Geophys. Res.* **105**, 15091–15106 (2000).

- Thomas, P. C. & Veverka, J. Downslope movement of material on Deimos. *Icarus* **42**, 234–250 (1980).
- Chapman, C. R. *et al.* Impact history of Eros: craters and boulders. *Icarus* (in the press).
- Hawkins, S. E. *et al.* Multispectral imager on the Near Earth Asteroid Rendezvous mission. *Space Sci. Rev.* **82**, 30–100 (1997).
- Veverka, J. *et al.* An overview of the NEAR multispectral imager-near-infrared spectrometer investigation. *J. Geophys. Res.* **102**, 23709–23727 (1997).

Acknowledgements

We thank the mission design, mission operations, and spacecraft teams of the NEAR Project at the Applied Physics Laboratory of Johns Hopkins University and the navigation team at the Jet Propulsion Laboratory for their efforts that made NEAR the first orbiter to make a successful landing. We are grateful for the helpful reviews provided by A. Rivkin and M. Cintala.

Correspondence and requests for materials should be addressed to J.V. (e-mail: jfv4@cornell.edu).

Shoemaker crater as the source of most ejecta blocks on the asteroid 433 Eros

P. C. Thomas*, J. Veverka*, M. S. Robinson† & S. Murchie‡

* Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA

† Department of Geological Sciences, Northwestern University, 309 Locy Hall, Evanston, Illinois 60208, USA

‡ Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

The loose material—regolith—on the surfaces of asteroids is thought to represent ballistically emplaced ejecta from impacts^{1,2} but the identification of source craters and the detailed study of the regolith modification have been hampered by the limited spatial resolution and area coverage of the few asteroids imaged by spacecraft. Here we report the results of global mapping of the asteroid 433 Eros from high-resolution images obtained by the NEAR-Shoemaker spacecraft. Based on the images and ejecta-emplacement models, we suggest that most large ejecta blocks on Eros originate from a relatively young 7.6-km-diameter crater. A large fraction of the ejecta from impacts pre-dating that crater has apparently been buried or eroded. The images also show evidence for the action of a variety of sorting environments for regolith particles after they are deposited on the surface.

To investigate the largest range of sizes of the Eros regolith we have mapped the mean horizontal dimensions of positive relief features (termed blocks, regardless of origin or size) larger than

15 m over the entire surface, and the sizes of blocks down to as small as 5 cm in local areas. There are 6,760 blocks more than 15 m across on the 1,125 km² of the asteroid. Nearly half (44%) of the volume of all blocks larger than 15 m, assuming heights are half the widths, is within a 7.6-km-wide crater (with the proposed name of Shoemaker). This region (Figs 1a, 1c and 2a) has a block volume per unit area 12 times that of the rest of Eros.

Blocks outside the Shoemaker crater are concentrated near the equator, and are relatively depleted in some longitudes (Fig. 2a). We have modelled trajectories of materials launched from the Shoemaker area as well as from other sites on Eros, using previously developed techniques³. Simulations involve launching particles at velocities between 0.5 and 11.5 m s⁻¹, 40° from the surface, radially from the centres of selected craters, and following them until impact. The accelerations are determined by gravity from a uniform density model of Eros⁴; launch velocities relative to the surface are added to local rotational vectors.

Trajectories from crater Shoemaker (even with variations from the nominal 40° launch angles) give patterns of impacts that approximate much of the observed pattern of blocks, whereas those arising from other large craters fail to approximate the pattern (Fig. 2b and c). Distinctive aspects of the observed distribution of large blocks are the latitudinal concentration and the clustering of blocks on the east side of crater Psyche combined with a dearth west of the centre of Psyche (90° W, Fig. 1b). The predicted decrease in blocks west of the middle of Psyche arises from the fact that Eros is concave on the 0–180° W side and convex on the other side. Ejecta from the Shoemaker crater approach the 90–180° W region from the west, and are largely shielded from this region by the long end of Eros. A source of blocks in Shoemaker may also explain why many blocks on the eastern slopes of Psyche (Fig. 1b, d) have not rolled downhill. Those launched from Shoemaker would impact this area with velocity vectors nearly normal to the surface. The prediction of

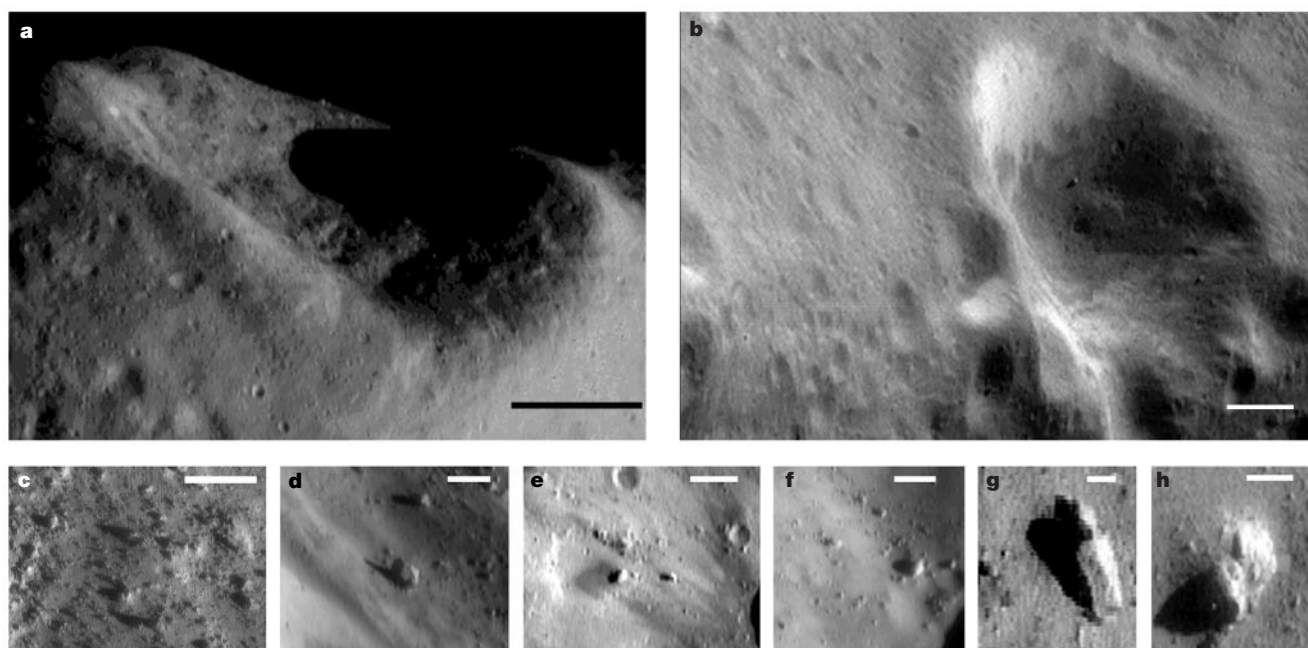


Figure 1 Large craters and blocks on Eros. **a**, Shoemaker crater viewed from the south. Mosaic of images 151027753, -815, -877 and -939. Scale bar is 2 km. **b**, The Psyche crater. Mosaic of images 130663215 and -481. View is from the west showing the 5.3-km crater which retains a subtle rim, has some large blocks on its eastern interior (facing wall of crater), but has very few blocks nearby to the west. Scale bar is 1 km. **c**, Interior of Shoemaker crater, showing high density of blocks. Image 156082111 at 27° S, 342° W. Scale bar is 200 m. **d**, Block resting on eastern slopes of crater Psyche. Image 139808668. Scale bar is 100 m. The interior of this crater is itself heavily cratered

and the blocks post-date Psyche. **e**, Blocks in a 1.5 km crater. Image 136594186, centred about 23° N, 14° W. Scale bar is 50 m. Blocks rest on degraded interior of this crater. **f**, Interior of crater Selene, the next largest crater on Eros after Psyche, about 3.2 km across, at 14° S, 14° W. Scale bar is 100 m. Image 140997355. These blocks also greatly post-date this crater. **g**, Angular block. Image 156082816 at 10.1° S, 5.3° W. Scale bar is 10 m. **h**, Degraded block. Image 155982204 at 4.0° S, 186.4° W. Scale bar is 20 m.

a relatively large number of blocks from Shoemaker on the eastern side of Himeros may be reconciled with the observed distribution because downslope motion in the area has clearly removed materials from the steeper slopes⁴. Significantly, ejecta from Psyche should be concentrated in the area that is instead observed to be relatively deficient in blocks (90°–150° W). This mismatch indicates that most large blocks from Psyche have been buried or eroded, or were possibly produced relatively less abundantly. Additionally, launches from craters Himeros (Fig. 2 and Fig. 1 of ref. 5) and Selene (15° S, 15° W) give predicted ejecta patterns very different in latitude and longitude concentrations from the observed pattern.

The observed distribution of blocks, the paucity of other large, fresh craters, and the lack of association of blocks with any other source craters suggest that a large fraction of blocks outside Shoemaker originated from that crater. Images show that most blocks in the larger (1–4 km) craters rest on degraded crater interiors, and therefore greatly post-date those craters (Fig. 1d–f).

The total calculated volume of blocks larger than 15 m is around $6 \times 10^7 \text{ m}^3$. The volume of material removed in the formation of the Shoemaker crater can be estimated to exceed 15 km^3 , or

$1.5 \times 10^{10} \text{ m}^3$, if the excavated volume is half that of the crater volume. The blocks are thus about 0.4% the excavated volume of Shoemaker crater, a value that may be compared to lunar observations that block masses can approach 5% of the total excavated from source craters⁶. The excavated volume of the 45 craters on Eros between 1 and 5.3 km in diameter is estimated to be approximately 12 km^3 , which is a more than adequate source volume for blocks that do not originate from Shoemaker.

Modelling of impacts into small objects predicts that the majority of ejecta from relatively large craters have low velocities, and thus return to the body⁷. The total excavated volume of craters over 1 km in diameter, including Shoemaker, is about 27 km^3 . Crater Himeros may have excavated in excess of 50 km^3 . If all of this material were retained on Eros, it could produce regolith around 75 m deep. Shoemaker crater alone could produce a global covering as much as 15 m in depth. Assuming that about three-quarters of Shoemaker crater's ejecta is retained⁷, but that it is not evenly distributed (Fig. 2), most of Eros would have accumulated a depth of a few metres up to less than 20 m of debris from Shoemaker crater.

Empirical relations of largest block size (L) and source crater diameter (D) have been developed^{8,9,10} for the Moon, Phobos and Deimos and asteroid Ida. The approximate relation is $L \approx 0.25D^{0.7}$ (where L, D are in metres¹⁰), or a 130-m block from a 7.6-km crater. Given that the largest block (150 m horizontal dimension) rests near the rim of Shoemaker, and there are only seven blocks larger than 100 m, the sizes of the blocks are consistent with a major contribution from Shoemaker crater.

The size distribution of blocks varies over Eros. The density of blocks, expressed as the cumulative number per unit area, is shown in Fig. 3. The global data have a cumulative log–log slope of -3.2 between sizes of 15 and 80 m. Cumulative slopes steeper than about -2 are consistent with very fragmented materials¹¹, steeper slopes

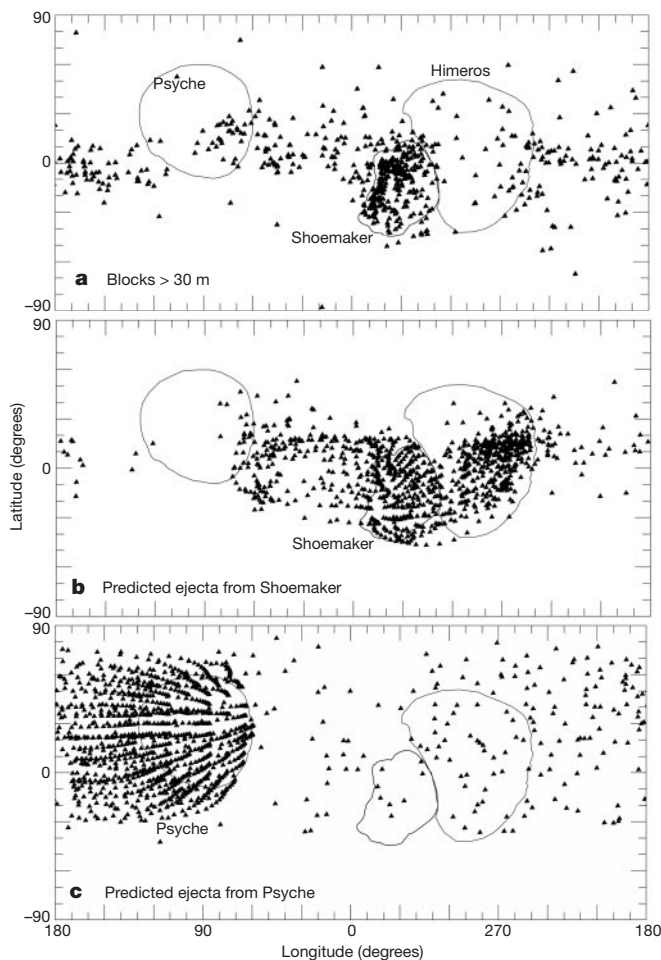


Figure 2 Distribution of blocks on Eros. **a**, Distribution of blocks larger than 30 m on Eros. The map is a simple cylindrical projection. The three largest craters are outlined; projection distorts relative areas owing to Eros' highly irregular shape. **b**, Predicted impact locations of model particles launched with velocities between 0.5 and 11.5 m s^{-1} , approximately 40° from a restored surface of Shoemaker crater. The pattern of ejecta remains similar for a wide range of ejection angles. The main properties of ejecta launched from Shoemaker are equatorial concentration and strong depletion in longitudes 100° – 150° W. The 0° – 180° W region is the concave side of the asteroid. **c**, Predicted impact locations of particles launched from Psyche crater.

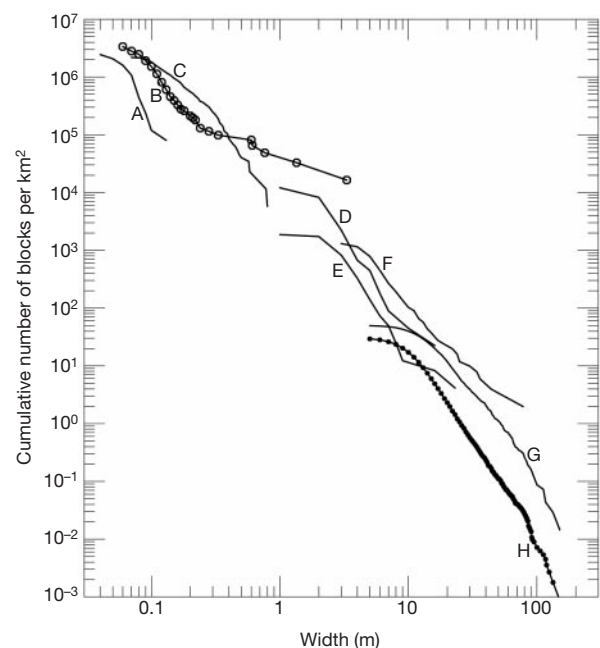


Figure 3 Block size frequencies on the surface of Eros. Cumulative numbers of blocks per unit area as a function of mean horizontal size. A–C indicate images 157417113, -178, and -198. D indicates image 155981791 at 2° S, 172° W. E indicates image 155883532 at 4° S, 19° W. F indicates global data. Curves show resolution effects at lower size fractions; digitization included many measurements in the marginally measurable sizes to ensure that all of the larger blocks were included. Curve H samples reliably to sizes of about 15 m. Because data for the small sizes cover only small areas and do not sample larger blocks, extrapolation from different curves can be misleading.

indicating greater fragmentation. (These surface densities of blocks have lower slopes, by around 0.2, than a volume density^{12,13}.) Studies of lunar blocks a few metres to tens of metres in size show cumulative slopes ranging from -1.8 to -3.7 (refs 6 and 14). Blocks on Phobos give cumulative slopes¹⁵ of -3.2. Thus, the cumulative sizes of Eros blocks as an average are consistent with blocks from craters on other airless objects. The slightly lower slope (-2.5 to -2.7) recorded for blocks within Shoemaker (curves F and G) suggests less processing of these materials than the global average, consistent with a relatively younger age. In the global data blocks larger than 80 m show a slope of -5. The very smallest blocks we mapped, less than 10 cm across, and found close to the NEAR landing site, show a slope of about -6. These steep slopes indicate strong depletion of the relatively larger blocks. The depletion mechanisms may be very different for 100-m blocks than for 20-cm blocks. Very high resolution images show metre-sized materials with a very low cumulative slope (curve B, white circles, Fig. 3). In this instance there is strong morphologic evidence of filling of areas around the larger blocks by a later deposition of fine material⁵.

Cumulative densities for a particular size vary by factors of over 20 at different locations. The differences may be due to the processes of ejection and reimpact, with additional size sorting caused by weathering, erosion and transport of materials. Processes that move materials on Eros may include gravity-driven flow, collapse into fractures¹⁶, thermal cycling¹⁷, electrostatics^{18,19}, and impact-related seismic effects^{20,21}.

The probable association of a large fraction of blocks on Eros with one large, fresh crater implies that a large fraction of blocks from older impacts has been buried or eroded, or that older craters produced relatively fewer ejecta. While it may be speculated that the formation of the Shoemaker crater on what amounts to a peak might influence aspects of crater excavation, there is as yet no firm evidence that the older craters produced relatively fewer blocks. Blocks on Eros have a great variety of shapes, some of which show crumbling of large pieces while others show small-scale erosion that generates fine aprons of debris around the base of blocks (Fig. 1g and h). The effectiveness of erosion of blocks is suggested by the insufficient depths of material deposited from the Shoemaker crater (average <15 m) to bury most of the larger blocks (>25 m) predicted to have been produced by the Psyche crater. The gradual elimination of blocks on the Moon by impact abrasion, disruption and burial has been shown to be an important component of lunar regolith processes²²⁻²⁴. Either erosion or burial of blocks is indicative of significant regolith depths on the asteroid. Erosion of blocks creates fine materials that contribute to the regolith, and burial of large blocks requires a regolith at least as deep as the smallest dimension of a boulder. □

Received 18 June; accepted 6 August 2001.

- Housen, K. R., Wilkening, L. L., Chapman, C. R. & Greenberg, R. Asteroidal regoliths. *Icarus* **39**, 317-351 (1979).
- Geissler, P. et al. Erosion and ejecta reaccretion on 243 Ida and its moon. *Icarus* **120**, 140-157 (1996).
- Thomas, P. C. Ejecta emplacement on the martian satellites. *Icarus* **131**, 78-106 (1998).
- Thomas, P. C. et al. Eros: Shape, slopes, and slope processes. *Bull. Am. Astron. Soc.* **32**, 994 (2000).
- Veverka, J. et al. The landing of the NEAR-Shoemaker spacecraft on asteroid 433 Eros. *Nature* **413**, 390-393 (2001).
- Cintala, M. J., Garvin, J. B. & Wetzel, S. J. The distribution of blocks around a fresh lunar mare crater. *Proc. Lunar Planet. Sci. Conf.* **13**, 100-101 (1982).
- Asphaug, E. & Melosh, H. J. The Stickney impact of Phobos: A dynamical model. *Icarus* **101**, 144-164 (1993).
- Moore, H. H. in *Analysis of Apollo 10 Photography and Visual Observations* 26-27 NASA SP-232 (NASA, Washington DC, 1971).
- Lee, S. W., Thomas, P. & Veverka, J. Phobos, Deimos, and the Moon: Size and distribution of crater ejecta blocks. *Icarus* **68**, 77-86 (1986).
- Lee, P. C. et al. Ejecta blocks on 243 Ida and on other asteroids. *Icarus* **120**, 87-105 (1996).
- Hartmann, W. K. Terrestrial, lunar, and interplanetary rock fragmentation. *Icarus* **10**, 201-213 (1969).
- Shoemaker, E. M. et al. Television observations from Surveyor V. JPL-NASA Technical Report 32-1246 (NASA, Washington DC, 1967).

- Meloy, T. & O'Keefe, J. A. Size distribution of lunar surface materials. *J. Geophys. Res.* **73**, 2299-2308 (1968).
- Cintala, M. J. & McBride, K. M. Block distributions on the lunar surface: A comparison between measurements obtained from surface and orbital photography. NASA Technical Memorandum 104804 (NASA, Washington DC, 1995).
- Thomas, P. C. et al. Phobos: regolith and ejecta blocks investigated with Mars Orbiter Camera images. *J. Geophys. Res.* **105**, 15091-15106 (2000).
- Horstmann, K. C. & Melosh, H. J. Drainage pits in cohesionless materials: Implications for the surface of Phobos. *J. Geophys. Res.* **94**, 12433-12441 (1989).
- Duennebie, F. K. & Sutton, G. H. Thermal moonquakes. *J. Geophys. Res.* **79**, 4351-4363 (1974).
- Lee, P. C. Dust levitation on asteroids. *Icarus* **124**, 181-194 (1996).
- Robinson, M. S., Thomas, P. C., Veverka, J., Murchie, S. & Carcich, B. The nature of ponded deposits on Eros. *Nature* **413**, 396-400 (2001).
- Houston, W. N., Moriawaki, Y. & Chang, C. S. Downslope movement of lunar soil and rock caused by meteoroid impact. *Proc. 4th Lunar Sci. Conf.* **3**, 2361-2364 (1973).
- Greenberg, R., Nolan, M. C., Bottke, W. F., Kolvoord, R. A. & Veverka, J. Collisional history of Gaspra. *Icarus* **107**, 84-97 (1994).
- Arvidson, R. E., Drozd, R. J., Hohenberg, C. M., Morgan, C. J. & Popeau, P. Horizontal transport of the regolith, modification of features, and erosion rates on the lunar surface. *Moon* **13**, 67-79 (1975).
- Horz, F., Schneider, E., Gault, D. E., Hartung, J. B. & Brownlee, D. Catastrophic rupture of lunar rocks: A Monte Carlo simulation. *Moon* **13**, 235-258 (1975).
- Reedy, R. C., Arnold, J. R. & Lal, D. Cosmic ray record in solar system matter. *Annu. Rev. Nucl. Part. Sci.* **33**, 505-537 (1983).

Acknowledgements

We thank K. May, B. Carcich, J. Joseph and R. Chomko for technical help. This work was funded by contracts from the National Aeronautics and Space Administration.

Correspondence and requests for materials should be addressed to P.C.T. (e-mail: thomas@cuspi.astro.cornell.edu).

The nature of ponded deposits on Eros

M. S. Robinson*, P. C. Thomas†, J. Veverka‡, S. Murchie‡ & B. Carcich†

* Department of Geological Sciences, Northwestern University, 1847 Sheridan Road, Evanston, Illinois, USA

† 310 Space Sciences Building, Cornell University, Ithaca, New York 14853, USA

‡ Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

One of the surprises of the NEAR-Shoemaker mission was that Eros's surface exhibits a wide variety of landforms, which are indicative of a global covering of loose fragmental debris¹. At one extreme in roughness is the Shoemaker Regio area, which is characterized by a high density of boulders up to 100 m across, slumps, slides, and finer blanketing material. At the other extreme are distinctive, flat deposits that appear smooth down to a resolution of 1.2 cm per pixel. Here we report the results of global mapping and colour analysis of these smooth deposits. They have formed most efficiently in restricted areas, and appear to be the result of deposition of finer material sorted from the upper portion of the asteroid's regolith. The smooth deposits constitute a family of features with a range of morphologies, but all appear to be the result of sedimentation. The geography of the deposits is consistent with some predicted aspects of photoelectric sorting, but these exotic transport and depositional mechanisms are not well understood. Deposits with the properties seen on Eros have no obvious analogues in previous lunar or asteroid data.

The margins of ponded deposits often sharply embay the sides of existing depressions (Figs 1 and 2). The surfaces of these smooth deposits are generally perpendicular to the local gravity gradient^{1,2}. Where smooth deposits occur within craters located on a regional slope they are offset from the geometric centre of the crater in the downslope direction¹. This relationship is consistent with emplacement with little to no shear strength, which allows the material to 'pond' to an equipotential surface. Such a depositional process

requires a mechanism for temporarily eliminating cohesive forces between particles. Ponded deposits can support steep-walled grooves, impact craters, and superposed boulders (Fig. 1), showing that subsequent to emplacement, compaction or another process has imparted non-zero shear strength to the deposit. In some places smooth materials do exhibit lobate margins where they are unconfin ed (Fig. 1) and such areas often do not exhibit the smoothness of the ponded deposits for reasons that remain unclear.

The ponds are not distributed randomly (Fig. 2). Although ponds occur at all latitudes, 231 (91%) of the 255 ponds over 30 m in diameter are located within $\pm 30^\circ$ of the equator. In particular there are zones within a few degrees of the equator centred west of the long ends of the asteroid that have distinctively more and larger ponded deposits. These are regions where the gravitational pull is lowest³. There is no geographical bias in the identification of the larger ponds (≥ 30 m per pixel). The best resolution coverage of the asteroid is in the equatorial regions; however, the large ponds plotted in Fig. 2 can be found in the 100-km orbit imaging coverage (~ 10 m per pixel), and much of the polar regions is well covered by

imaging from 50-km orbits (~ 5 m per pixel). At small scales (≤ 10 m in diameter), it is not known if ponds are geographically restricted, but we note that of the ponds < 30 m in diameter, only 30% lie within 60° of either pole.

The diameters of host craters can be used to give an upper estimate of the depth of their ponds, by assuming an unmodified crater shape similar to that of fresh craters (depth/diameter ratio is 0.2) found elsewhere on the asteroid. Such an estimate represents an upper bound, because most of the craters have experienced some degree of degradation. The maximum calculated depths of the ponded deposits are only a few metres. If these volumes have been obtained from the craters' interiors, then the ponds could have formed by the redistribution of less than an average depth of 10 cm of material from the slopes of the depressions (assuming comparable density). Ponds larger than 30 m in diameter have not been found in craters over 1 km in diameter. This may be due to the mass movement of debris and boulders in the floors of the larger craters that masks, or interferes with, the formation of ponded deposits (Fig. 2). In addition, we have not identified ponded

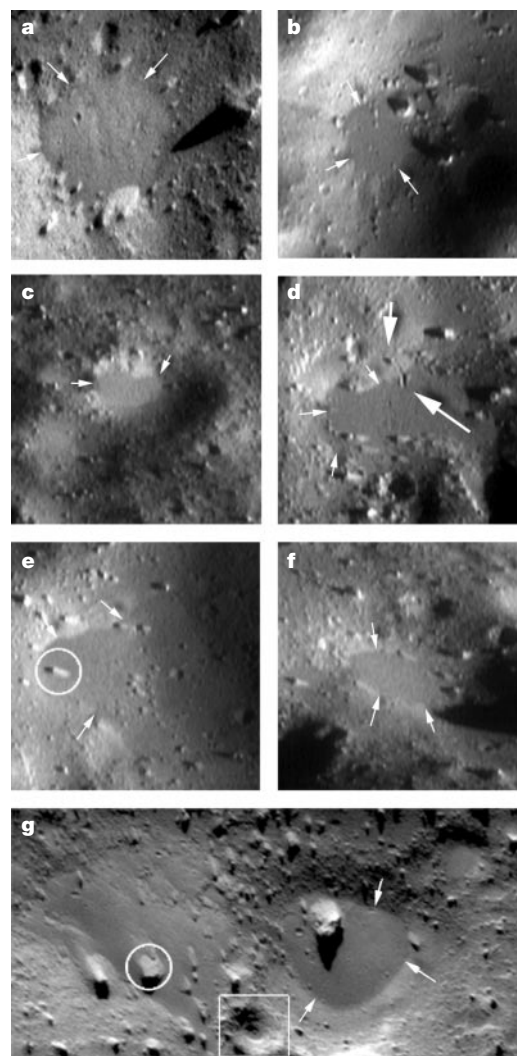


Figure 1 Views of ponded deposits. The material that forms the ponded deposits sharply embays pre-existing topography (small arrows) and in some cases is also seen to drape and pile up against pre-existing topography (circles). The large, long arrow in **d** indicates a groove in the pond, showing that pond material develops shear strength after deposition; the large, short arrow points out the low-angle impactor that formed the groove. The square box in **g** highlights a topographically isolated crater with a small interior pond, showing that the source of ponded material did not flow in from surrounding

terrain. **a**, MET image 15588731, 2.8° S, 183.9° W, 76 m in diameter, 0.6 m per pixel. **b**, MET image 147932871, 0.2° N, 182.1° W, 210 m in diameter, 3.1 m per pixel. **c**, MET image 154460965, 2.1° S, 169.4° W, 22 m in diameter, 1.6 m per pixel. **d**, MET image 154461005, 2.2° S, 170.0° W, 190 m in diameter, 1.6 m per pixel. **e**, MET image 154461045, 3.48° S, 170.84° W, 190 m in diameter, 1.6 m per pixel. **f**, MET image 154460865, 3.86° S, 166.51° W, 139 m in diameter, 1.7 m per pixel. **g**, MET images 155888584–8605, 2.4° S, 179.1° W, 90 m in diameter, 0.5 m per pixel.

deposits in craters that show obvious evidence of downslope movement of regolith.

Ponded material was observed in the final descent images⁴, and appears smooth in images of about 1.2 cm per pixel. We hence infer an upper bound on the grain sizes of ponded material of less than a centimetre. In contrast, surrounding non-pond surfaces are rich in boulders tens of centimetres to metres in diameter. The fine fraction of the regolith that composes the ponded deposits could have been formed by primary spallation of original bedrock or by disaggregation of larger components of the regolith through thermal cycling, small-scale impacts, or abrasion during slide and slump events.

High-spatial-resolution colour observations of the asteroid show that many of the ponds have colour properties unambiguously distinct from the surrounding terrain (Fig. 3); in the visible range the ponds are relatively blue (high 550-nm/760-nm reflectance ratios) and have a deeper one-micrometre mafic band due to absorption by olivine and pyroxene (as indicated by their 950-nm/760-nm ratios). The ponds also exhibit an elevated reflectance at 760 nm relative to their surroundings (<5%). Until now, all remote-sensing measurements of the asteroid imply a compositionally homogeneous body^{5,6}, discounting the possibility that the ponds are a different rock type. Alternative explanations for these colour differences include secondary alteration (space weathering), mineral sorting, or grain size and packing^{6,7}; these possibilities are discussed below.

As fresh crystalline material in the lunar regolith is altered by exposure to the space environment over time (space weathering), it becomes darker and redder and the mafic absorption is diminished^{8,9}. This optical alteration occurs as submicroscopic metallic iron is vapour-deposited on grains; the vapour results from micrometeorite impact and solar wind sputtering^{10,11}. Also important is the formation of relatively large aggregates of quenched melt and adhering regolith particles, called agglutinates, which tend to darken lunar soils. However, there is considerable debate over the rates at which space weathering (including agglutinate production) takes place in asteroid regoliths^{10–12}. In the case of lunar-like space weathering, the ponded material would correspond to less altered material (relatively immature), and surrounding materials would be more altered (relatively mature). To be consistent with space-weathering trends, either the ponded material is substantially younger than the material it embays or a fine, relatively

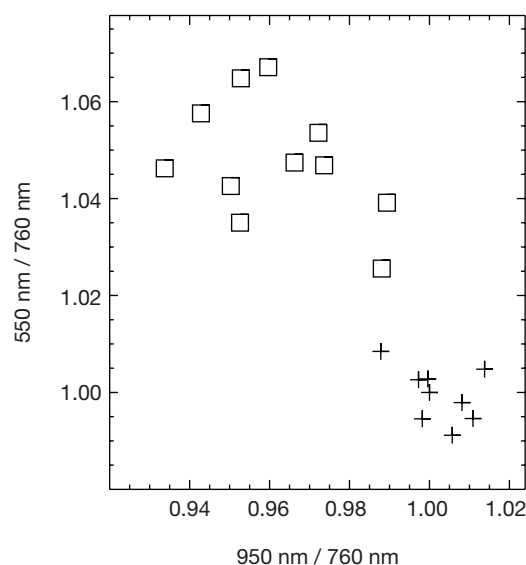


Figure 3 Colour properties of ponded deposits. Relative colour ratios for ponds (squares) and typical inter-pond regions (crosses) on Eros from colour sequence acquired on 15 January 2001 are normalized to the median of the inter-pond values (MET images 155006469–155008009).

unaltered, crystalline component of the regolith is being locally sorted and collected to form the ponds (after initial deposition).

An alternative explanation to space weathering is that a sorting process is preferentially removing silicate material (olivines and pyroxenes) from a silicate plus metal regolith and depositing it in the ponds. Another possibility is that the pond's colour properties are simply due to grain size differences. It is well known that changes in grain size and packing can greatly affect the colour properties of silicate materials¹³. For mafic rocks, as grain size decreases, colour becomes redder at visible wavelengths. However, at very fine grain sizes ($\leq 50 \mu\text{m}$) some mafic materials reverse this trend and their visible colour becomes bluer¹³. To explain the colour properties of the ponds through differences in grain size, extremely fine grain sizes ($\ll 50 \mu\text{m}$) would be required. Such fine particles could be

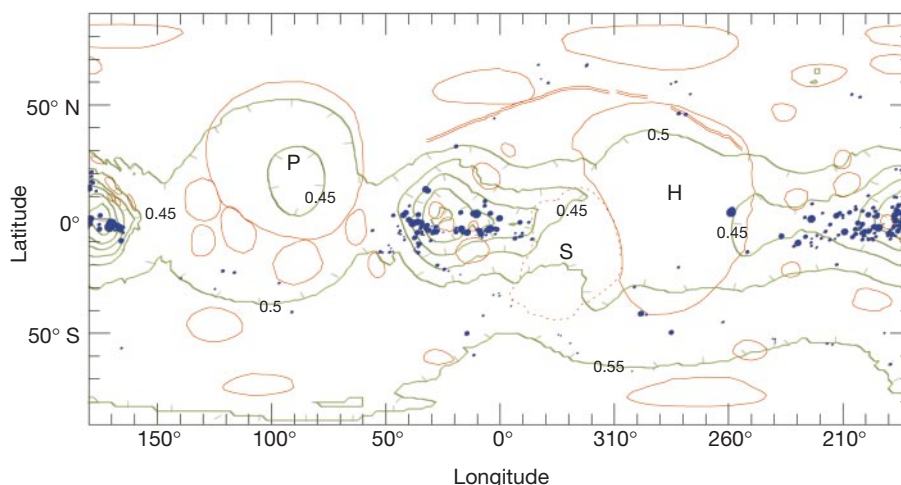


Figure 2 Map of pond distributions and gravity on Eros. Smooth ponded deposits (blue) are found preferentially in the equatorial belt near the long ends of the asteroid, sizes of symbols are proportional to pond sizes. Ponds are not found within craters greater than 1 km in diameter (red) nor in areas of relatively high gravity (green contours labelled in black). Higher gravity and steep slopes, such as we found in the interiors of the Psyche (P),

Shoemaker (S) and Himeros (H) craters, may be the primary cause of the longitudinal restrictions. The value of interior closed contour near 170° W longitude is 0.25 cm s^{-2} , and of the interior closed contour found near 20° W longitude is 0.30 cm s^{-2} (contour interval is 0.05 cm s^{-2}).

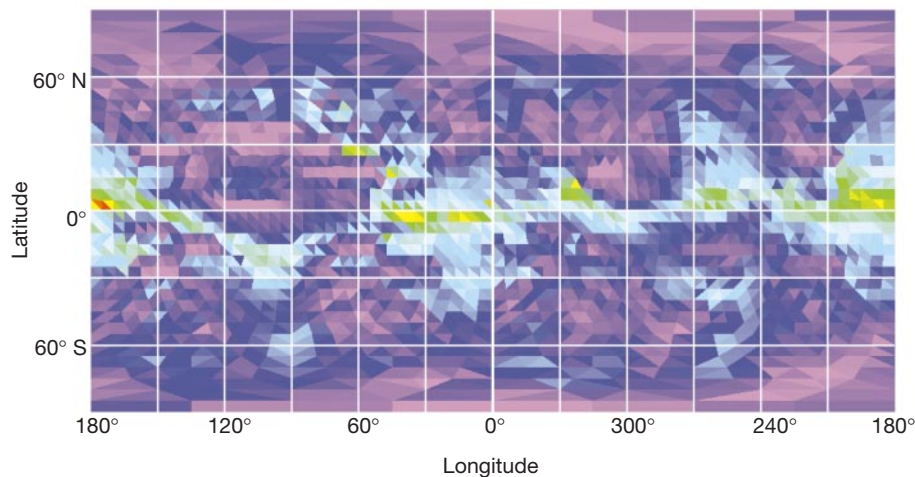


Figure 4 Relative amount of time an area receives illumination between 85° and 90° incidence angle. Red, yellow, green, indigo and violet represent decreasing time at the terminator over an erosian year, respectively (red to violet represents more than a fourfold

decrease). Enhanced electrostatic environment due to extremes in time spent at the terminator may be an important latitudinal control of the ponded deposits.

particularly sensitive to the electrostatic effects discussed below.

Formation of the ponded deposits requires a mechanism to separate the fine fraction from coarser components of the regolith, and a mechanism to concentrate the fine particles in depressions, most efficiently along the equatorial belt. Terrestrial sorting mechanisms typically include aeolian, aqueous, glacial, or gravitational processes. The asteroidal sorting environment is probably limited to gravitational processes (slides, slumping, seismic shaking, ejecta flows) and exotic processes such as electrostatic levitation¹⁴ or thermal creep^{15,16}.

That the larger ponds form in the equatorial belt may be an observational key to understanding their origin. The spatial distributions of large (>15 m) ejecta blocks are consistent with ejection primarily from the Shoemaker crater¹⁷; by inference, finer ejecta from this event ought also to be concentrated at equatorial latitudes. Stratigraphically, the amount of ejecta from this relatively recent, large impact would mask any previously existing pond forms. These relationships imply that the materials that form the ponded deposits are no older than this large impact. A possible scenario is that the fine component of Shoemaker ejecta was somehow sorted and moved independently from the coarser material during emplacement, and possibly further concentrated during subsequent seismic shaking following later impact events. This is an attractively simple explanation, but it does not adequately explain the detailed geographic restrictions of pond distribution. Alternatively, electrostatic effects could result in the preferential removal of very fine material as charge differentials build up along the terminator. Observations of fine suspended particles near the lunar surface associated with terminator crossings¹⁸ have led to models that yield photoelectric charge differentials between illuminated and dark areas sufficient to levitate particles up to tens of micrometres in size¹⁴. The critical element for achieving sufficient charging is the duration over which local boundaries between illuminated and shadowed areas remain stable, and is basically proportional to times over which high local incidence angles prevail¹⁴. Eros's extreme obliquity (88°) means that the equatorial belt of the asteroid receives incident illumination within 5° of the terminator for a significantly larger percentage of the year than do other regions. Modelling of a full Eros year shows that local regions of particularly long terminator durations correspond well with the mapped distribution of larger ponds (Fig. 4). Exceptions occur in areas of steeper slopes, and relatively higher local gravity (Fig. 2). Calculated depths of material sorted locally from the regolith also

show sharply greater values within 10° of the equator, suggesting that whatever process is involved is more efficient very close to the equator. There is also an excellent correspondence of pond locations with the areas having the lowest gravity, that is, the sites most favourable for small particle levitation.

Although electrostatic effects will preferentially liberate the fine size component of the regolith, deposition in discrete, flat deposits presumably requires gravitational transport of particles still charged sufficiently to have no shear strength thus allowing them to flow into depressions. Once transported into depressions, changes in illumination conditions would lead to decreased charging efficiency, allowing the particles to form consolidated deposits which would, as observed, display evidence of finite shear strength (rocks sitting on ponds, collapse pits, craters, grooves, and so on). The colour properties of the ponds may be explained by a simple grain size effect alone, through a separation of immature material from a more mature background, or even concentration of silicate-rich material from silicate plus metallic iron regolith. In any case, electrostatic levitation and its ability to sort out the finer component of the regolith may explain the extreme smoothness and the distinctive colour of the ponded deposits. □

Received 18 June; accepted 8 August 2001.

1. Veverka, J. *et al.* Imaging of small-scale features on 433 Eros from NEAR: Evidence for a complex regolith. *Science* **292**, 484–488 (2001).
2. Cheng, A. F. *et al.* Laser altimetry of small-scale features on 433 Eros from NEAR-Shoemaker. *Science* **292**, 488–491 (2001).
3. Thomas, P. C. *et al.* Eros: Shape, topography and slope processes. *Icarus* (in the press).
4. Veverka, J. *et al.* The landing of the NEAR-Shoemaker spacecraft on asteroid 433 Eros. *Nature* **413**, 390–393 (2001).
5. Bell, J. F. III *et al.* Near-IR reflectance spectroscopy of 433 Eros from the NIS instrument on the NEAR mission. 1. Low-phase angle observations. *Icarus* (in the press).
6. Murchie, S. *et al.* Color variations on Eros from NEAR multispectral imaging. *Icarus* (in the press).
7. Clark, B. E. *et al.* NEAR infrared spectrometer photometry of asteroid 433 Eros. *Icarus* (in the press).
8. McCord, T. B. & Johnson, T. V. Lunar spectral reflectivity (0.30 to 2.50 microns) and implications for remote mineralogical analysis. *Science* **169**, 855–858 (1970).
9. McCord, T. B. & Adams, J. B. Progress in remote optical analysis of lunar surface composition. *Moon* **7**, 453–474 (1973).
10. Hapke, B. Space weathering from Mercury to the asteroid belt. *J. Geophys. Res.* **106**, 10039–10073 (2001).
11. Pieters, C. M. *et al.* Space weathering on airless bodies: Resolving a mystery with lunar samples. *Meteor. Planet. Sci.* **35**, 1101–1107 (2000).
12. Matson, D. L., Johnson, T. V. & Veeder, G. J. Soil maturity and planetary regoliths: the Moon, Mercury, and the asteroids. *Proc. Lunar Sci. Conf.* **8**, 1001–1011 (1977).
13. Adams, J. B. & Felice, A. L. Spectral reflectance 0.4 to 2.0 microns of silicate rock powders. *J. Geophys. Res.* **72**, 5705–5715 (1967).
14. Lee, P. Dust levitation on asteroids. *Icarus* **124**, 181–194 (1996).
15. Duennebie, F. K. & Sutton, G. H. Thermal moonquakes. *J. Geophys. Res.* **79**, 4351–4363 (1974).

16. Lee, P. C. *Physical Properties and Processing of Asteroid Regoliths and Interiors*. Thesis, Cornell Univ. (1977).
17. Thomas, P. C., Veverka, J., Robinson, M. S. & Murchie, S. Shoemaker crater as the source of most ejecta blocks on the asteroid 433 Eros. *Nature* **413**, 394–396 (2001).
18. Criswell, D. R. in *Photon and Particle Interactions with Surfaces in Space* (ed. Grard, R. J. L.) 545–556 (Reidel, Dordrecht, 1973).

Acknowledgements

We thank the mission design, mission operations, and spacecraft teams of the NEAR Project at the Applied Physics Laboratory of Johns Hopkins University and the navigation team at the Jet Propulsion Laboratory for their efforts that resulted in making NEAR Shoemaker the first spacecraft to orbit and land on an asteroid. The NEAR-Shoemaker mission and this work were made possible by NASA. We thank A. Rivkin and M. Cintala for suggestions.

Correspondence and requests for materials should be addressed to M.S.R. (e-mail: robinson@earth.northwestern.edu).

Experimental long-lived entanglement of two macroscopic objects

Brian Julsgaard, Alexander Kozhekin & Eugene S. Polzik

Institute of Physics and Astronomy, University of Aarhus, 8000 Aarhus, Denmark

Entanglement is considered to be one of the most profound features of quantum mechanics^{1,2}. An entangled state of a system consisting of two subsystems cannot be described as a product of the quantum states of the two subsystems^{3–6}. In this sense, the entangled system is considered inseparable and non-local. It is generally believed that entanglement is usually manifest in systems consisting of a small number of microscopic particles. Here we demonstrate experimentally the entanglement of two macroscopic objects, each consisting of a caesium gas sample containing about 10^{12} atoms. Entanglement is generated via interaction of the samples with a pulse of light, which performs a non-local Bell measurement on the collective spins of the samples⁷. The entangled spin-state can be maintained for 0.5 milliseconds. Besides being of fundamental interest, we expect the robust and long-lived entanglement of material objects demonstrated here to be useful in quantum information processing, including teleportation^{8–10} of quantum states of matter and quantum memory.

In 1935 Einstein, Podolsky and Rosen (EPR) formulated¹ what they perceived as a paradox created by quantum mechanics. Since then, the EPR correlations and other types of entanglement have been extensively analysed, notably by Bell². Entangled or inseparable states are fundamental to the field of quantum information, specifically to quantum teleportation of discrete^{8,9} and continuous¹⁰ variables and to quantum dense coding of discrete¹¹ and continuous^{12,13} variables, to name a few examples. The majority of experiments on entanglement until now deal with entangled states of light^{3,4,8–11,14}. Entangled states of discrete photonic variables (spin-half systems)^{3,4} as well as entangled states of continuous variables (quadrature-phase operators) of the electro-magnetic field¹⁴ have been generated experimentally. Entangled states of material particles are much more difficult to generate experimentally; however, such states are vital for storage and processing of quantum information. Recently, entangled states of four trapped ions have been produced¹⁵, and two atoms have been entangled via interaction with a microwave photon field¹⁶.

Here we describe an experiment on the generation of entangle-

ment between two separate samples of atoms containing 10^{12} atoms each, along the lines of a recent proposal⁷. Not only do we demonstrate a quantum entanglement at the level of macroscopic objects, our experiment proves feasible a new approach to the quantum interface between light and atoms suggested in refs 7, 17. It is a step towards the other protocols proposed^{7,17}, such as the teleportation of atomic states and quantum memory. The entanglement is generated through a non-local Bell measurement on the two samples' spins, performed by transmitting a pulse of light through the sample.

The ideal EPR entangled state of two sub-systems described by continuous non-commuting variables $\hat{X}_{1,2}$ and $\hat{P}_{1,2}$ —such as the positions and momenta of two particles, for example—is the state for which $\hat{X}_1 + \hat{X}_2 \rightarrow 0$, $\hat{P}_1 - \hat{P}_2 \rightarrow 0$. Recently^{5,6}, the necessary and sufficient condition for the entanglement or inseparability for such gaussian quantum variables has been cast in the form of an inequality involving only the variances of variables: $\langle (\delta(\hat{X}_1 + \hat{X}_2))^2 \rangle + \langle (\delta(\hat{P}_1 - \hat{P}_2))^2 \rangle < 2$. In our experiment, the quantum variables that are analogous to the position and momentum operators are two projections of the collective spin (total angular momentum) of an atomic sample. The analogy is evident from the commutation relation $[\hat{J}_x, \hat{J}_y] = i\hat{J}_z$, which can be rewritten as $[\hat{X}, \hat{P}] = i$, where $\hat{X} = \hat{J}_x/\sqrt{J_x}$ and $\hat{P} = \hat{J}_y/\sqrt{J_x}$ if the atomic sample is spin-polarized along the x axis with $\hat{J}_x$ having a large classical value J_x . For two spin-polarized atomic samples with $J_{x1} = -J_{x2} = J_x$ the above entanglement condition translates into

$$\delta J_{12}^2 \equiv \delta J_{z12}^2 + \delta J_{y12}^2 < 2J_x \quad (1)$$

where we introduce the notations $\delta J_{z12}^2 = \langle (\delta(\hat{J}_{z1} + \hat{J}_{z2}))^2 \rangle$ and $\delta J_{y12}^2 = \langle (\delta(\hat{J}_{y1} + \hat{J}_{y2}))^2 \rangle$. The interpretation of condition (1) comes from the recognition of the fact that for both atomic samples in coherent spin states (CSS) the equality $\delta J_{y,z}^2 = J_x/2$ holds, that is, the inequality (1) becomes the equality. Entanglement between atoms of the two samples is, therefore, according to condition (1), equivalent to spin variances that are smaller than in samples in a CSS^{18,19} characterized by uncorrelated individual atoms. The entangled state of this type is a two-mode 'squeezed' state for the continuous spin variables^{5,7}. A spin squeezed state of a single macroscopic atomic ensemble has been generated previously^{21,22}.

Entanglement is produced via interaction of atoms with polarized light. A polarized pulse of light is described by Stokes operators obeying the same commutation relation as spin operators $[\hat{S}_y, \hat{S}_z] = i\hat{S}_x$. $\hat{S}_x$ is the difference between photon numbers in x and y linear polarizations, $\hat{S}_y$ is the difference between polarizations at $\pm 45^\circ$, and $\hat{S}_z$ is the difference between the left- and right-hand circular polarizations along the propagation direction, z . In our experiment light is linearly polarized along the x axis. Hence the two pairs of continuous quantum variables engaged in the entanglement protocol are $\hat{J}_z$ and $\hat{J}_y$ for atoms and $\hat{S}_z$ and $\hat{S}_y$ for light.

Here we report on the generation of a state of two separate caesium gas samples (Fig. 1), which obey the entanglement condition (1). As shown in refs 7 and 17, when an off-resonant pulse is transmitted through two atomic samples with opposite mean spins $J_{x1} = -J_{x2} = J_x$, the light and atomic variables evolve as

$$\begin{aligned} \hat{S}_y^{\text{out}} &= \hat{S}_y^{\text{in}} + \alpha \hat{J}_{z12}, \hat{S}_z^{\text{out}} = \hat{S}_z^{\text{in}} \\ \hat{J}_{y1}^{\text{out}} &= \hat{J}_{y1}^{\text{in}} + \beta \hat{S}_z^{\text{in}}, \hat{J}_{y2}^{\text{out}} = \hat{J}_{y2}^{\text{in}} - \beta \hat{S}_z^{\text{in}}, \hat{J}_{z1}^{\text{out}} = \hat{J}_{z1}^{\text{in}}, \hat{J}_{z2}^{\text{out}} = \hat{J}_{z2}^{\text{in}} \end{aligned} \quad (2)$$

where α and β are constants. The first line describes the Faraday effect (polarization rotation of the probe). The second line shows the back action of light on atoms, that is, spin rotation due to the angular momentum of light. According to equations (2), the measurement of $\hat{S}_y^{\text{out}}$ reveals the value of $\hat{J}_{z12} = \hat{J}_{z1} + \hat{J}_{z2}$ (provided α is large enough, so that $\hat{S}_y^{\text{in}}$ is relatively small) without changing this value. It follows from equations (2) that the total y spin projection for both samples is also conserved, $\hat{J}_{y1}^{\text{out}} + \hat{J}_{y2}^{\text{out}} = \hat{J}_{y1}^{\text{in}} + \hat{J}_{y2}^{\text{in}}$. The procedure can be repeated with another pulse of

light measuring the sum of y components, $\hat{J}_{y1} + \hat{J}_{y2}$, again in a non-demolition way, while at the same time leaving the previously measured value of $\hat{J}_{z1} + \hat{J}_{z2}$ intact. As a result, the sum of the z components and the sum of the y components of the spins of the two samples are known exactly in the ideal case, and therefore the two samples are entangled according to condition (1), since the uncertainties on the left-hand side become negligible.

An important modification of the above protocol is the addition of a magnetic field oriented along the direction x , which allows us to use a single entangling pulse to measure both z and y spin projections, as described in the Methods section.

The schematic of the experimental set-up is shown in Fig. 1. The two cells are coated from inside with a paraffin coating which enhances the ground-state coherence time T_2 up to 5–30 ms depending on the density of atoms. $|J_x|$ and T_2 are measured by the magneto-optical resonance method (ref. 20 and references therein).

After the samples are prepared in CSS, as described in the legend to Fig. 1, optical pumping is switched off and a probe pulse is sent through. Its Stokes operator S_y is measured by a polarizing beam splitter with two balanced detectors. The differential photocurrent from the detectors is split in two and its $\cos(\Omega t)$ and $\sin(\Omega t)$ power spectral components $(S_{y\cos}^{\text{out}}(\Omega))^2$ and $(S_{y\sin}^{\text{out}}(\Omega))^2$ are measured by lock-in amplifiers. By repeating this sequence many times we obtain the variances for these components, in short, spectral variances, which according to equation (3) in the Methods section are $(\delta S_{y\cos}^{\text{out}}(\Omega))^2 = (\delta S_{y\sin}^{\text{out}}(\Omega))^2 + 1/2\alpha^2\delta J_{z12}^2 \equiv 1/2\delta S^2 + \kappa\delta J_{z12}^2$, where $\kappa = \frac{1}{2}\alpha^2$, and similarly for $(\delta S_{y\sin}^{\text{out}}(\Omega))^2$ with substitution $J_z \rightarrow J_y$. The coefficient $\alpha = \sigma\gamma n/4FA\Delta \approx 2.5$ is estimated¹⁷ using $\sigma \approx \lambda^2/2\pi$ as the resonant dipole cross-section, $\gamma = 5$ MHz as the full width of the optical transition, $F = 4$ as total angular momentum of the hyperfine ground state, $A = 2$ cm² as the probe beam cross-section,

and $n = 10^{13}$ as the number of photons in the 0.45 ms probe pulse with power 5 mW.

The spectral variance data, $\Delta = (\delta S_{y\cos}^{\text{out}}(\Omega))^2 + (\delta S_{y\sin}^{\text{out}}(\Omega))^2 = \delta S^2 + \kappa\delta J_{z12}^2 + \kappa\delta J_{y12}^2 \equiv \delta S^2 + \kappa\delta J_{12}^2$, is plotted in Fig. 2. The linear fit to it, $\Delta(J_x)$, is the quantum limit of noise corresponding to the coherent spin state of samples (see Fig. 2 legend for details).

Using $\Delta(J_x)$ as a reference level we can now devise a measurement procedure, which will verify the presence of an entangled state. Namely, if for a certain state of the two ensembles the spectral variance of the signal $\Delta_{\text{EPR}} = \delta S^2 + \kappa\delta J_{\text{EPR}}^2$ obeys the inequality:

$$\Delta_{\text{EPR}} = \delta S^2 + \kappa\delta J_{\text{EPR}}^2 < \delta S^2 + 2\kappa J_x = \Delta(J_x) \quad (4)$$

then apparently $\delta J_{\text{EPR}}^2 < 2J_x$ holds for such a state and therefore this state is entangled in accordance with condition (1). It is, of course, necessary to use otherwise identical conditions for measurements of Δ_{EPR} and $\Delta(J_x)$, that is, the same value of $2J_x$ and the same probe intensity and detuning in order to keep the constant κ unchanged throughout the entire experiment. As mentioned above, the value of $2J_x$ has been controlled by a magneto-optical resonance measurement to better than 5%. In the actual experiment described below the measurements of Δ_{EPR} and $\Delta(J_x)$ have been conducted in turn at identical conditions at the repetition rate of 500 Hz.

The measurement sequence aimed at the generation and verification of the entanglement consists of the optical pumping pulses, which induces a CSS, the entangling pulse, which induces an entangled state (pulse I) in the samples, and the verifying pulse, which occurs after a delay time τ and verifies the entanglement (pulse II). These pulses have the same duration and optical frequency as the probe pulse used for the CSS measurements. Between the two pulses the joint spin state of the two samples is subject to decoherence. The photocurrents from the two pulses are subtracted electronically and the variance of the difference, Δ_{EPR} , is

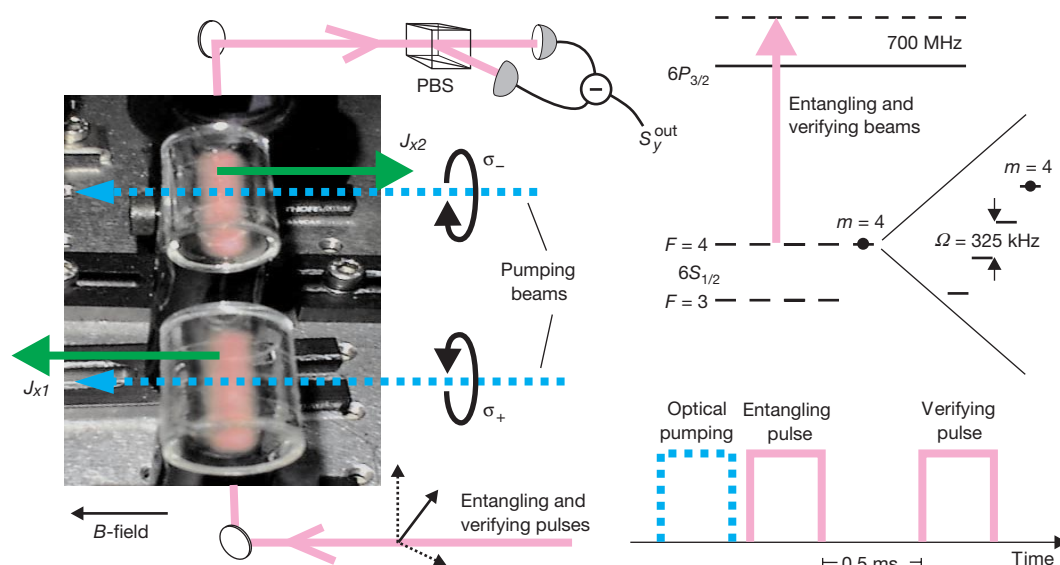


Figure 1 The experimental set-up, atomic level structure and the sequence of optical pulses. Fluorescence of atoms interacting with light is seen inside the cells (artificial colour). The two atomic samples in glass cells 3×3 cm at approximately room temperature are placed in a highly homogenous magnetic field of 0.9 G ($\Omega = 325$ kHz) surrounded by a magnetic shield. Caesium atoms are optically pumped into $F = 4$, $m_F = 4$ ground state in the first cell and into $F = 4$, $m_F = -4$ in the second cell to form coherent spin states oriented along the x axis for cell 1 and along $-x$ for cell 2. Optical pumping is achieved with circularly polarized 0.45 ms pulses at 852 nm and 894 nm with opposite helicity for the two cells. The mean spin value for each sample is given by $J_x = N\sum_{m=-4}^4 m\rho_{mm} = 4Np \rightarrow 4N$, where N is the number of atoms and ρ_{mm} is the probability of an atom being in the m state, with the limiting value corresponding

to a perfect spin polarization, $p = 1$. The $|J_x|$ for the two cells is adjusted to be equal to or better than 5%. After the optical pumping is completed 0.45 ms long entangling and verifying pulses separated by a 0.5-ms delay are sent through. Both pulses are blue-detuned by $\Delta = 700$ MHz from the closest hyperfine component of the D_2 line at 852 nm. The thermal atomic motion is not an obstacle in this experiment but is rather helpful. The probe beam covers most but not the whole volume of the atomic sample, but the duration of the pulses is longer than the transient time of an atom across the cell. Therefore each light pulse effectively interacts with and entangles all atoms of the samples. In addition, a large detuning of the probe makes the Doppler broadening insignificant.

measured as discussed in the Methods section. The vanishing Δ_{EPR} corresponds to two repeated measurements on the total spin state of the two samples producing the same results, that is, it corresponds to a perfect knowledge of both $\hat{J}_{z12}$ and $\hat{J}_{y12}$ and therefore to a perfectly entangled state. In the experiment the minimal value of Δ_{EPR} is $2\delta S^2$ owing to the quantum noise of the entangling and verifying pulses.

The results of measurements with $\tau = 0.5$ ms are shown in Fig. 3. The results are normalized to the CSS limit $\Delta(J_x)$ (the linear fit in Fig. 2). This limit thus corresponds to the unity level in Fig. 3. The raw experimental data $\Delta_{\text{EPR}}/\Delta(J_x)$ for the entangled state are shown as stars. The values below the unity level verify that the entangled state of the two atomic samples has been generated and maintained for 0.5 ms. The derivation of the degree of entanglement from the data in Fig. 3 is given in the Methods section. The degree of entanglement calculated operationally from the data without additional assumptions is $\xi = (35 \pm 7)\%$. The degree of entanglement useful for teleportation that was calculated using an additional, experimentally proved assumption of the initially CSS for both samples is higher, $\xi' = (52 \pm 7)\%$. The predicted teleportation fidelity performed with the same two pulses as used in the present paper (see Methods section) is $\mathcal{F} = 55\%$, which is above the classical limit of 50%. The factors limiting the fidelity are of a technical nature, and we expect higher fidelity to be well within reach.

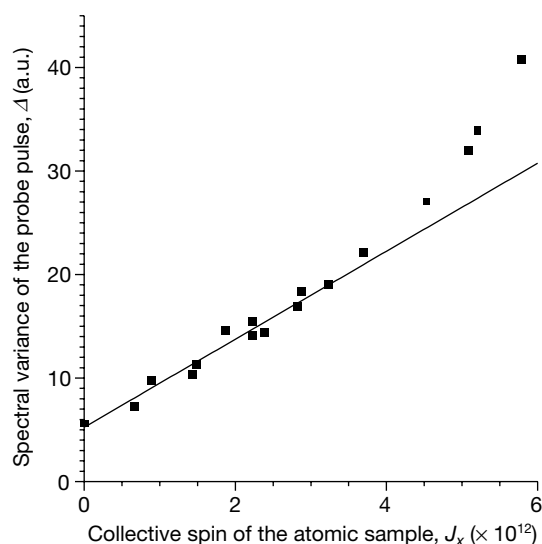


Figure 2 Determination of the coherent spin state limit for entanglement. The total measured variance of the two quadratures of the photocurrent $\Delta = (\delta S_{\text{yos}}^{\text{out}}(\Omega))^2 + (\delta S_{\text{ysin}}^{\text{out}}(\Omega))^2 = \delta S^2 + \kappa \delta J_{z12}^2 + \kappa \delta J_{y12}^2 \equiv \delta S^2 + \kappa \delta J_{12}^2$ is plotted as a function of J_x . J_x is measured independently by magneto-optical resonance method and is varied by heating the cells and by adjusting optical pumping. In the absence of atoms in the cells the measured spectral variance is due to the initial probe state variance δS^2 . δS^2 is at the vacuum (shot) noise level, which has been verified experimentally by checking its characteristic linear dependence on the probe power. With atoms present the measured variance grows linearly with J_x at low densities, which proves that there is no classical contribution to the spin noise and that therefore the observed atomic fluctuations at these densities are entirely due to quantum atomic noise²¹. The degree of the spin polarization for the data in the figure (shown as squares) is nearly perfect, $p \geq 95\%$, which means that the spin state is very close to the coherent spin state (CSS). We therefore conclude that a linear fit to the observed data corresponds to the CSS of both atomic samples for which $\delta J_{12}^2 = 2J_x$ and hence this line, which we denoted by $\Delta(J_x)$, establishes the noise level corresponding to the right-hand side of the inequality (4). Deviations of the observed variance from the linear fit at high atomic numbers are due to the nonlinearly growing contribution of classical technical noise of the spin state because of the technical noise of lasers, the non-ideal cancellation of the back action of the probe on the two samples, and so on.

The imperfect entanglement comes from several factors. First, the vacuum noise of the entangling pulse prohibits a perfect preparation of the spin state, especially for a small number of atoms. Second, the deviation of the initial spin state from CSS, which is due to the classical noise of the lasers, becomes more pronounced as the number of atoms grows. Finally, losses of light on the way from one cell to another, as well as the spin-state decay between the two measurements preclude perfect entanglement at all atomic densities. The spin-state decay is caused by collisions and the quadratic Zeeman effect.

Measurements with delay times longer than 0.8 ms demonstrate no entanglement. The decoherence process can be illustrated by visualizing two ellipses in the phase space, corresponding to the two samples, with the size along the long axes of the order of three, in units of the coherent state (experimentally measured value). When the ellipses start to dephase (rotate) the total noise along the short axis becomes equal to the coherent state noise (zero degree of entanglement) after the time of the order of 1.2 ms if the initial entanglement of 65% (maximal degree allowed by the experimental light noise) is assumed. After a 0.6-ms delay a degree of entanglement of 35% should be expected, given the experimentally measured dephasing time of 5 ms. These numbers agree reasonably well with observations.

It is instructive to analyse the difference between the degree of entanglement and the degree of classical correlations. For uncorrelated atomic samples the normalized variance of the difference between the two photocurrents would be equal to two, using the notation of Fig. 3. The pure atomic part of this variance V_{uncorr} would be approximately 1.5 for medium atomic densities. For the actual data in Fig. 3 the atomic part of the variance is approximately $V_{\text{corr}} = 0.25$. The degree of correlation, $1 - V_{\text{corr}}/V_{\text{uncorr}}$, is 83% for this example. This number is much higher than the degree of

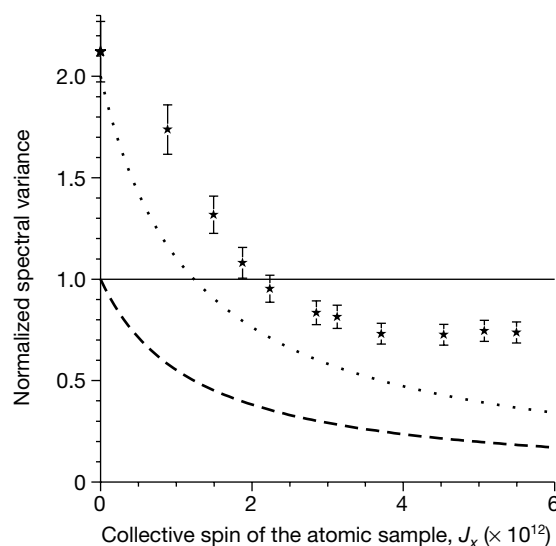


Figure 3 Demonstration of the entangled spin state for two atomic samples. The results are plotted as a function of J_x and are normalized to the CSS limit $\Delta(J_x)$ (the linear fit in Fig. 2). This limit—the boundary between entangled (below the line) and separable states—thus corresponds to the unity level in the figure (solid line). The raw experimental data $\Delta_{\text{EPR}}/\Delta(J_x)$ —data for the entangled spin state, which has lived for $\tau = 0.5$ ms—are shown as stars. The values below the unity level verify that the entangled state of the two atomic samples has been generated and maintained for 0.5 ms. The minimal possible level for $\Delta_{\text{EPR}}/\Delta(J_x)$ (maximum possible entanglement) is equal to $2\delta S^2/\Delta(J_x)$ (dotted line), that is, it is set by the total quantum noise of the two pulses. The normalized shot noise level of the verifying pulse, $\delta S^2/\Delta(J_x)$ (dashed line), which is used for calculations of the degree of entanglement, is also shown.

entanglement, implying that entanglement requires something stronger than classical correlations. The reason for this difference is that in quantum mechanics the measuring device (light in this case) becomes entangled with the measured object and therefore the noise of these two subsystems cannot be treated independently.

Thus, we have demonstrated on-demand generation of entanglement of two separate macroscopic objects, which can be maintained for more than 0.5 ms. The state we have demonstrated is not a maximally entangled ‘Schrödinger’s cat’ state, which for 10^{12} atoms would not survive even for a femtosecond under the conditions of our experiment. Our state is similar to a two-mode squeezed state, and is an example of a non-maximally entangled state which is suitable for a particular purpose, for example, atomic teleportation. The long lifetime of this multi-particle entanglement is due to a high symmetry of the generated state. Entanglement manifests itself only in the collective properties of the two ensembles. Therefore a loss of coherence for a single atom makes a negligible effect on the entanglement, unlike in a maximally entangled multi-particle state. The entanglement is generated by means of light propagating through the two samples and therefore the samples can be distant, as is required for communications. The off-resonant character of the interaction used for the creation of entanglement allows for the potential extension of this method to other media, possibly including solid-state samples with long-lived spin states. $\square$

Methods

Entangling spin components with a single light pulse

The Larmor precession of the J_z, J_y components with a common frequency Ω does not change their mutual orientation and size and therefore does not affect the entanglement. On the other hand, the precession allows us to extract information about both z and y components from a single probe pulse as described below. Moreover, in the laboratory frame the spin state is now encoded at the frequency Ω , and as usual an a.c. measurement is easier to reduce to the quantum noise level than a d.c. measurement. Measurements of the light noise can now be conducted only around its Ω spectral component. By choosing a suitable radio-frequency value for Ω we can reduce the probe noise $S_{yz}^{\text{in}}(\Omega)$ to the minimal level of the vacuum (shot) noise. In the presence of the magnetic field the spin behaviour is described by the following equations: $\dot{J}_z(t) = \Omega \hat{J}_y(t), \dot{J}_y(t) = -\Omega \hat{J}_z(t) + \beta \hat{S}_y(t)$, whereas the Stokes operators still evolve according to equation (2). Solving the spin equations and using equation (2) we obtain

$$\hat{S}_y^{\text{out}}(t) = \hat{S}_y^{\text{in}}(t) + \alpha [\hat{J}_{z1} \cos(\Omega t) + \hat{J}_{y1} \sin(\Omega t)] \quad (3)$$

The $\hat{J}_{x,y}$ components are now defined in the frame rotating with the frequency Ω around the magnetic field direction x . It is clear from equation (3) that by measuring the $\cos(\Omega t)/\sin(\Omega t)$ component of $\hat{S}_y^{\text{out}}(t)$ we can acquire the knowledge of the z/y spin projections. Simultaneous measurement of both spin components of the two atomic samples is possible because they commute: $[\hat{J}_{z1} + \hat{J}_{z2}, \hat{J}_{y1} + \hat{J}_{y2}] = i(\hat{J}_{x1} + \hat{J}_{x2}) = 0$.

Degree of entanglement

The differential noise for the entangling (pulse I) and verifying (pulse II) pulses can be expressed using equation (3) as:

$$\begin{aligned} \Delta_{\text{EPR}} &= \delta S_{\text{cos}}^2 + \delta S_{\text{sin}}^2 = (\delta S_{y\text{cos}}^{\text{out}}(\Omega) - \delta S_{y\text{cos}}^{\text{in}}(\Omega))^2 + (\delta S_{y\text{sin}}^{\text{out}}(\Omega) - \delta S_{y\text{sin}}^{\text{in}}(\Omega))^2 \\ &= \delta S_{\text{in}}^2 + \kappa \left\{ \kappa^{-1} \delta S_{\text{in}}^2 + [\delta(f_{z12})_1 - \delta(f_{z12})_2]^2 + [\delta(f_{y12})_1 - \delta(f_{y12})_2]^2 \right\} \\ &\equiv \delta S_{\text{in}}^2 + \kappa \delta f_{\text{EPR}}^2 \end{aligned} \quad (5)$$

The variance δf_{EPR}^2 determines the uncertainty of the spin components of the state prepared by the entangling pulse and measured after time τ . It contains two contributions; δS_{in}^2 , due to the quantum noise of the entangling pulse, and all other terms in the curly brackets, which are due to decoherence during the time τ . The verifying pulse in our experiment is also not noiseless and therefore its variance, δS_{in}^2 , also contributes to Δ_{EPR} . In the experiment $\delta S_{\text{in}}^2 = \delta S_{\text{in}}^2 = \delta S^2$. The exact entanglement condition, $\Delta_{\text{EPR}} < \Delta(J_x)$, is obtained by substituting equation (5) into equation (4). The degree of entanglement can be defined as $\xi = 1 - (\delta f_{\text{EPR}}^2/2J_x) = 1 - (\Delta_{\text{EPR}} - \delta S^2)/(\Delta(J_x) - \delta S^2)$. ξ varies from 0 (separable state) to 1 (perfect entanglement). The highest degree of entanglement calculated operationally from the data is $\xi = (35 \pm 7)\%$.

An alternative calculation of the degree of entanglement can be carried out⁷, which takes into account that the initial state of both samples is characterized by the CSS noise level. Equation (2) of ref. 7, using the notation of the present paper, yields for the highest possible degree of entanglement (defined again as $\xi = 1 - (\delta f_{\text{EPR}}^2/2J_x)$) and created by the first pulse in our protocol, the value $\xi_{\text{theory}} = 1 - \eta_{\text{theory}}$ where the degree of the ‘‘two-mode squeezing’’ is, according to ref. 7, $\eta_{\text{theory}} = \delta S^2/(\delta S^2 + 2\kappa J_x) = \delta S^2/\Delta(J_x)$. This is

the best entanglement possible for a given ratio of the light noise and the atomic spin noise contribution. This value of entanglement would correspond to the minimal possible difference between the two pulses, $\Delta_{\text{EPR}}^{\text{min}} = 2\delta S^2$. We note that this value is closer to unity (stronger entanglement) than the degree of entanglement $\xi = 1 - (\Delta_{\text{EPR}}^{\text{min}} - \delta S^2)/(\Delta(J_x) - \delta S^2) = 1 - \delta S^2/(\Delta(J_x) - \delta S^2)$ calculated in the previous paragraph. This is because the entanglement in ref. 7 is calculated assuming that the samples prior to entanglement are known to be in the CSS. This fact has been experimentally verified in the present paper for up to intermediate atomic densities, so we may use this way of calculating the degree of entanglement at these densities as well. However, we are interested in the degree of entanglement, which has survived the delay time and which has been measured by the verifying pulse. For various reasons, including the decoherence during the delay time, the differential noise between the two pulses does not reduce to its minimal value, $\Delta_{\text{EPR}} > \Delta_{\text{EPR}}^{\text{min}} = 2\delta S^2$. Therefore the actual degree of entanglement, as witnessed by the measurement, is lower than the maximum possible: $\xi' = 1 - \eta_{\text{exper}} < \xi_{\text{theory}}$. Here the degree of squeezing, $\eta_{\text{exper}} = (\Delta_{\text{EPR}} - \delta S^2)/\Delta(J_x) > \eta_{\text{theory}}$ obtained from the data in Fig. 3 is worse than the theoretical best value, η_{theory} , owing in part to the diffusion of the state during the delay time between the entangling and the verifying pulses. From Fig. 3 we find $\eta_{\text{exper}} = 0.48$ and therefore $\xi' = 52\%$ for $J_x \approx 3.5 \times 10^{12}$.

Expected fidelity of teleportation

It is also possible to estimate the fidelity of teleportation which would be achieved if the second, verifying pulse were (instead of verification) used for the Bell measurement on one of the entangled samples and a sample to-be-teleported. Using the value of ξ' and equation (3) from ref. 7 we obtain a value for the fidelity of teleportation of $\mathcal{F} = 55\%$ for $J_x \approx 3.5 \times 10^{12}$. This is higher than the classical boundary of $\mathcal{F} = 50\%$.

We note that the entangled state reported here has a random element in it, namely, every (ideal) entangling measurement creates the state $J_{z1} + J_{z2} = x, J_{y1} + J_{y2} = p$ with x, p random measured values. However, this state is as efficient for teleportation, for example, as is a state $J_{z1} + J_{z2} = 0, J_{y1} + J_{y2} = 0$.

Received 11 June; accepted 27 July 2001.

1. Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
2. Bell, J. S. *Speakable and Unsayable in Quantum Mechanics* (Cambridge Univ. Press, Cambridge, 1988).
3. Aspect, A., Dalibard, J. & Roger, G. Experimental test of Bell’s inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804–1807 (1982).
4. Torgerson, J. R., Branning, D., Monken, C. H. & Mandel, L. Experimental demonstration of the violation of local realism without Bell inequalities. *Phys. Lett. A* **204**, 323–328 (1995).
5. Duan, Lu, M., Giedke, G., Cirac, J. I. & Zoller, P. Inseparability criterion for continuous variable systems. *Phys. Rev. Lett.* **84**, 2722–2725 (2000).
6. Simon, R. Peres-Horodecki separability criterion for continuous variable systems. *Phys. Rev. Lett.* **84**, 2726–2729 (2000).
7. Duan, L. M., Cirac, J. I., Zoller, P. & Polzik, E. S. *Phys. Rev. Lett.* **85**, 5643–5646 (2000).
8. Bouwmeester, D., Pan, J. W., Mattle, K., Eibl, M., Weinfurter, H. & Zeilinger, A. Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
9. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
10. Furusawa, A. et al. Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
11. Mattle, K., Weinfurter, H., Kwiat, P. G. & Zeilinger, A. Dense coding in experimental quantum communication. *Phys. Rev. Lett.* **76**, 4656–4659 (1996).
12. Braunstein, S. L. & Kimble, H. J. Dense coding for continuous variables. *Phys. Rev. A* **61**, 042302 (2000).
13. Ban, M. Quantum dense coding via a two-mode squeezed-vacuum state. *J. Opt. B* **1**, L9–L11 (1999).
14. Ou, Z. Y., Pereira, S. F., Kimble, H. J. & Peng, K. C. Realization of the Einstein-Podolsky-Rosen paradox for continuous variables. *Phys. Rev. Lett.* **68**, 3663–3666 (1992).
15. Sackett, C. A. et al. Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
16. Rauschenbeutel, A. et al. Step-by-step engineered multiparticle entanglement. *Science* **288**, 2024–2028 (2000).
17. Kuzmich, A. & Polzik, E. S. Atomic quantum state teleportation and swapping. *Phys. Rev. Lett.* **85**, 5639–5642 (2000).
18. Kitagawa, M. & Ueda, M. Spin squeezed states. *Phys. Rev. A* **47**, 5138–5143 (1993).
19. Wineland, D. J., Bollinger, J. J., Itano, W. M., Moore, F. L. & Jeomzem, D. J. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A* **46**, R6797–R6800 (1992).
20. Kupriyanov, D. V. & Sokolov, I. M. Optical detection of magnetic resonance by classical and squeezed light. *Quant. Opt.* **4**, 55–70 (1992).
21. Hald, J., Sørensen, J. L., Schori, C. & Polzik, E. S. Spin squeezed atoms: a macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1322 (1999).
22. Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum nondemolition measurement. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).

Acknowledgements

We gratefully acknowledge the contributions of J. Hald, J. L. Sørensen, C. Schori and A. Verchovsky. We also thank I. Cirac, A. Kuzmich, A. Sørensen and P. Zoller for discussions.

Correspondence and requests for materials should be addressed to E.P. (e-mail: polzik@ifa.au.dk).

A Bragg glass phase in the vortex lattice of a type II superconductor

T. Klein*, I. Joumard*, S. Blanchard*, J. Marcus*, R. Cubitt†, T. Giamarchi‡ & P. Le Doussal§

* Laboratoire d'Etudes des Propriétés Electroniques des Solides—CNRS, BP166, 38042 Grenoble Cedex 9, France

† Institut Laue Langevin, BP 156, 38042 Grenoble Cedex 9, France

‡ Laboratoire de Physique des Solides, CNRS-UMR8502, UPS Bât. 510, 91405 Orsay, France

§ CNRS-Laboratoire de Physique Théorique de l'Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, France

Although crystals are usually quite stable, they are sensitive to a disordered environment: even an infinitesimal amount of impurities can lead to the destruction of crystalline order¹. The resulting state of matter has been a long-standing puzzle. Until recently it was believed to be an amorphous state in which the crystal would break into 'crystallites'². But a different theory³ predicts the existence of a novel phase of matter: the so-called Bragg glass, which is a glass and yet nearly as ordered as a perfect crystal. The 'lattice' of vortices that contain magnetic flux in type II superconductors provide a good system to investigate these ideas⁴. Here we show that neutron-diffraction data of the vortex lattice provides unambiguous evidence for a weak, power-law decay of the crystalline order characteristic of a Bragg glass. The theory also predicts accurately the electrical transport properties of superconductors; it naturally explains the observed phase transitions^{4–6} and the dramatic jumps in the critical current^{7,8} associated with the melting of the Bragg glass. Moreover, the model explains experiments as diverse as X-ray scattering in disordered liquid crystals^{9,10} and the conductivity of electronic crystals^{11,12}.

After the Bragg glass was first proposed³, its existence was supported by further analytical^{13–15} and numerical^{16,17} calculations. However, up to now experimental evidence for this phase has been indirect. For instance, the presence of a Bragg glass existing as a stable thermodynamic phase is expected to impose a universal phase diagram for type II superconductors^{3,18,19}—and similar experimental phase diagrams have indeed been observed in a large variety of compounds^{4–6}. Moreover, recent transport measurements in BiSrCaCuO (ref. 20) are consistent with the Bragg glass predictions^{21,3}. But given the complexity of these materials, it is quite difficult to rule out other interpretations of the data. To decide unambiguously on the nature of the vortex solid, it is necessary to use an imaging technique able to probe directly the topology of the system. Even if decoration and imaging experiments show large regions free of dislocations²² or well ordered structures when the lattice is set into motion^{23,24}, these regions are still too small to provide an unambiguous signature of the Bragg glass: the expected power-law divergence of the Bragg peaks. Diffraction experiments such as small angle neutron scattering (SANS)^{25–28} can be performed at larger magnetic fields, and can thus bring the scale at which disorder becomes relevant down to the experimental resolution. When properly analysed, SANS experiments are thus an ideal method for deciding on this issue.

The experiments we report here were performed on a large, single-phase (K,Ba)BiO₃ crystal (mass, 300 mg; superconducting transition temperature T_c , 23 K). This compound has the advantage of being isotropic, enabling us to avoid complications due to anisotropy such as a possible crossover from three dimensions to two dimensions (3D–2D). Our experiments were performed on the D11 line at the Institut Laue Langevin in Grenoble. Neutrons are coherently diffracted by the flux lines when the Bragg condition is

fulfilled. In order to study the positional order of the vortex lattice, it is necessary to measure the correlation functions perpendicular to the vortices (that is, in the detector plane). However, the poor experimental in-plane resolution (a few lattice constants) does not allow us to probe these correlations directly. The solution is to measure the correlations along the vortex lines which can, on the contrary, be probed with good accuracy and are directly related to the in-plane correlations through the elastic constants. The sample is rocked through the Bragg angle (by an angle ω) and the diffracted intensity is integrated in the detector plane, giving the so-called rocking curve. The scattering geometry is shown in the top panel of Fig. 1. In the vicinity of a wavevector $\mathbf{K}_0$, the integrated intensity is then related to the structure factor $S(q)$ through: $I(\omega) = F^2 \int dq_x \int dq_y S(q_x, q_y, K_0 \omega)$, where F is the standard form factor of a single vortex²⁹. The rocking curves obtained at 2 K for $K_0 \approx 3^{1/2}/2.2\pi a_0$, where a_0 is the vortex lattice spacing, are shown in the lower panel of Fig. 1. The most direct way to check for the existence of the Bragg glass would be to deconvolute the experimental resolution and directly check the predicted³ power-law decrease of $S(q)$. In superconductors, the intensity in the tails of the rocking curve is too weak to do this reliably. Fortunately, as we will show, this power-law decay can also be obtained from the magnetic-field dependence of the rocking curves, which turns out to provide unambiguous evidence for the existence of the Bragg glass.

As shown in Fig. 1, $I/F^2 a_0$ is field-independent below a field $H^* \approx 0.7$ T but rapidly collapses at higher fields. However, despite this strong reduction of the intensity at the Bragg angle (that is, of $S(q_z = 0)$), the half-width at half-maximum of the rocking curves, σ , remains at about 0.18° , independent of the magnetic field; this value is close to the experimental resolution and remains constant

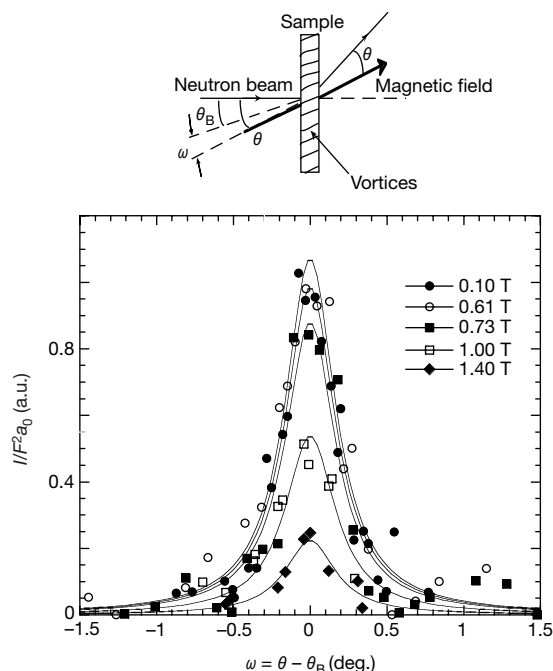


Figure 1 Angular dependence of the diffracted neutron intensity (the rocking curve) in a (K,Ba)BiO₃ single crystal, and the sample geometry. Data were taken at 2 K for different magnetic fields. The neutron beam (of wavelength 10 Å) was approximately parallel to the [111] direction of the crystal. The sample was cooled in a slightly oscillating field (of the order of 5% of the d.c. field) in order to obtain a good in-plane orientation order. ω is the angle by which the sample is rocked away from the Bragg angle θ_B (θ being the angle between the neutron beam and the magnetic field—upper panel for the scattering geometry). The solid lines are lorentzian-distribution fits to the data with constant width at half-maximum $\sigma \approx 0.18^\circ$ (similar fits can be obtained using gaussian distributions). Those data show that the diffracted intensity (rocking curves) collapse without any broadening above ~ 0.7 T.

up the highest magnetic field. This is a striking result, because the London³⁰ theory predicts that the area under the rocking curve (Fig. 1) should remain roughly constant. A similar decrease of the diffracted intensity without any broadening of the rocking curve was also observed²⁵ in BiSrCaCuO, and attributed to a 3D–2D crossover. This explanation, cannot, however, hold in the isotropic (K,Ba)BiO₃ system.

How can this unusual behaviour be understood? We first recall that for a perfect crystal, the finite experimental resolution (in momentum, $\Delta q \approx 1/\xi$) would transform the ideal delta-function peak of the structure factor at K_0 into a broadened peak of height ξ^d and half-width $1/\xi$, (where d is the spatial dimension). If the positional order in the system is not perfect but decays over the characteristic length scale R_a (for which displacements are of the order of the lattice spacing (that is, $u(R_a) \approx a_0$)), the observed Bragg peak remains determined by the experimental resolution as long as $\xi < R_a$. However, the structure factor will become dependent on the decay of the translational order in the system for $\xi > R_a$. This will then allow us to distinguish between an amorphous system full of dislocations and the Bragg glass. In the first case, the translational order decays very rapidly beyond R_a (for example, exponentially with the distance), leading to a peak of height R_a^d and half-width in momentum of $1/R_a$ (the influence of the experimental resolution is negligible in this case); a decrease in R_a will thus lead to a decrease of the height of the peak and a broadening of the peak. The situation is very different in the Bragg glass. Indeed, in this case the positional order decays only weakly beyond R_a (as a power law of the distance with an exponent $\eta \approx 1$). The same power-law decay is predicted³ both perpendicular and parallel to the vortices, and the corresponding characteristic lengths R_{az} and R_{axy} are connected through $R_{az} = R_{axy} \sqrt{c_{44}/c_{66}}$ (where the c_{ij} are elastic constants). In the absence of any experimental limitation this would lead³ to power-law divergent Bragg peaks for $q \rightarrow 0$: $S(q) \approx 1/q^{d-\eta}$. In a real experiment, however, ξ suppresses this divergence and the shape of the peak is then controlled both by R_a (in the appropriate direction) and ξ .

Our calculation shows that, for $\xi > R_a$, the structure factor (and hence the rocking curve) has the form shown in Fig. 2. The most important point is that, for $d \geq \eta$, the width at half-maximum of

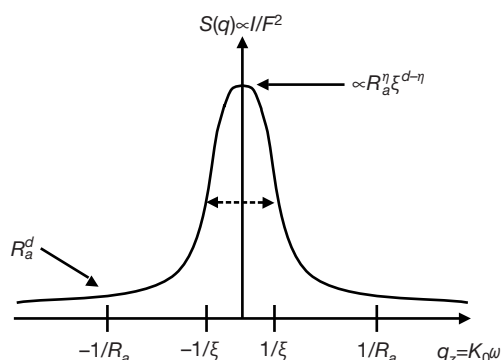


Figure 2 Bragg-glass predictions for the angular dependence of the diffracted neutron intensity. In this phase, the positional order is expected to decay with an exponent η beyond some characteristic length scale R_a . ξ is the experimental resolution. The arrows indicate the values of $S(q)$ for $q = 0$ and $1/R_a$ respectively. If the disorder increases (R_a decreases), the height of the peak decreases as R_a^η but the peak does not broaden as the half-width at half-maximum is always given by the experimental resolution $1/\xi$ and not $1/R_a$. Such behaviour can only be observed if $\eta \approx d$ (the dimensionality of the system), and is thus intrinsically related to the weak decay of the correlation functions. The height of the peaks gives a direct measure of the characteristic length R_a . To make the connection with Fig. 1, one should set $d = 1$ owing to the integration of the diffracted intensity in the detector plane, and q should be replaced by q_z . As done in Fig. 1, because $q_z = K_0 \omega$, an extra normalization factor a_0 is necessary to ensure that $K_0 \int I(\omega) d\omega = \int S(q) dq$ (note that the experimental resolution is constant in angle).

the peak remains entirely given by the experimental resolution $1/\xi$ for both $R_a > \xi$ and $R_a < \xi$, whereas the height of the peak $S(q_z = 0)$ is proportional to R_a^η for $R_a < \xi$. In the Bragg glass, a decrease in R_a thus leads to a decrease of the diffraction peak without any broadening. As shown in Fig. 1, this behaviour is in excellent agreement with the experimental data for $H > H^* \approx 0.7$ T (for $H < H^*$, $R_a > \xi$ and the intensity remaining constant). We note that such an unusual collapse of the rocking curve is intrinsically related to the power-law behaviour of $S(q)$, and could not be observed for any faster decay of the positional order. Moreover, because the rocking curves measure the diffraction along the z direction, we have to set $d = 1$ in the above expressions and this behaviour thus imposes $\eta \approx 1 = d$ (the half-width at half-maximum σ is again expected to increase as a_0/R_a for $(\eta - d) \log(\xi/R_a) > 1$), in excellent agreement with the Bragg-glass model³ ($\eta \approx 1$ – 1.2) as well as numerical simulations¹⁷. The collapse of the rocking curve without any broadening is thus direct proof of the existence of the Bragg glass.

In order to test further the consistency of this explanation, we have extracted the positional correlation length R_{az} from the data, using the Bragg-glass formula for the intensity ($I(\omega = 0)/F^2 \approx R_{az}$). The unknown overall normalization of the intensity is fixed by the condition that $R_{az} \approx \xi = 1/(K_0 \sigma) \approx 50a_0$ when the intensity starts to depend on the magnetic field. The field dependence of R_{az} is plotted in Fig. 3. Because, for $H \gg H_{c1}$ (where H_{c1} is the lower critical field), increasing the field is equivalent to increasing the effective disorder in the system, the observed decrease of $I(\omega = 0)/F^2$ is in good agreement with the expected decrease of R_{az} . The solid line in Fig. 3 corresponds to a $1/B^{5/2}$ variation for R_{az} , which is consistent with the calculated dependence of R_{az} on the magnetic field B using a simple elastic description for the vortex lattice¹⁹. The intensity finally drops rapidly towards zero at a field B_m lying close to the onset of the second peak in magnetization measurements, thus confirming the theoretical interpretation of the second peak in terms of Bragg glass ‘melting’. Moreover, neutron data give a very reasonable value for the positional correlation length at the transition: R_{az} (here of the order of R_{axy}) $\approx 20a_0$. As

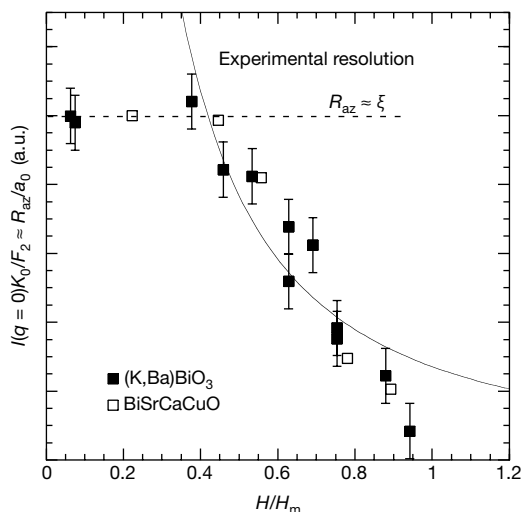


Figure 3 Magnetic-field dependence of the positional length R_{az} for (K,Ba)BiO₃ and BiSrCaCuO (ref. 24) crystals. In the Bragg-glass model, R_{az} is directly related to the maximum of the diffracted intensity (intensity at the Bragg angle): $R_a \approx I(\omega = 0)/F^2$. The x axis has been renormalized to the field H_m above which the diffracted intensity disappears. The solid line corresponds to a $1/B^2$ dependence for R_{az} as expected from a classical elastic theory. The horizontal dotted line is the limit above which R_{az} becomes larger than the experimental resolution $\xi \approx 50a_0$ in (K,Ba)BiO₃. The intensity drops rapidly towards zero at H_m , probably reflecting the proliferation of dislocations at the order–disorder transition.

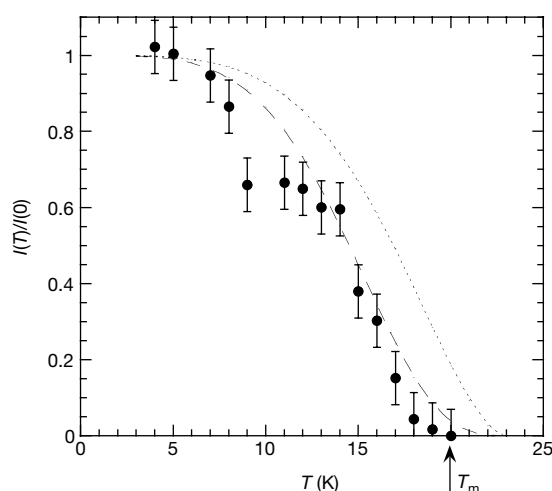


Figure 4 Temperature dependence of the maximum of the diffracted neutron intensity in a (K,Ba)BiO₃ crystal. The intensity (I) at the Bragg angle has been normalized to its value at $T = 0$ K. The dotted line is the $1/\lambda^4$ dependence expected in the London theory. The data rapidly deviate from this simple dependence. The dashed line is the Bragg-glass

shown in Fig. 3, this interpretation can also be applied to BiSrCaCuO. Indeed, similar behaviour is observed in the two systems provided that the field axis is renormalized by the field for which the diffracted intensity vanishes (this moreover shows that there is only one relevant parameter: B_m , which is directly related to $R_a(0)$, that is, the residual amount of disorder in the system).

The Bragg-glass theory also explains the puzzling behaviour of the measured temperature dependence of the intensity I (Fig. 4). As shown, the $1/\lambda^4$ dependence of the intensity expected in the standard London³⁰ model ($I \approx F^2$, with F defined previously) provides very poor agreement with the experimental data, using a classical two-fluid model for temperature dependence of the magnetic penetration length $\lambda(T)$ (dotted line). This discrepancy was also found to exist for YBaCuO (ref. 26) and remained unexplained. The neutron data can be fitted by introducing a phenomenological modification of the temperature dependence of the penetration length. However, such a strong deviation from the classical model has not been observed in the $\lambda(T)$ deduced from muon spin resonance relaxation data²⁶. This shows that, besides being an *ad hoc* fit to the data, such a modification of $\lambda(T)$ does not in fact contain any physical meaning. As muons are mostly sensitive to the behaviour of individual vortices, the deviation observed in neutron data can be directly related to the long-range correlation in the vortex lattice.

Indeed, in the Bragg-glass model, an additional temperature dependence arises from $R_a(T)$ that is expected to vary as $(L_c/\xi_0^2)^{2/3}$ in the elastic theory (where ξ_0 is the superconducting coherence length and L_c the collective pinning length that describes the strength of pinning). Assuming that pinning is related to fluctuations in the transition temperature (the so-called δT_c pinning model), $L_c \propto \xi_0^{2/3}$, and the resulting temperature dependence is plotted in Fig. 4 (dashed line). As shown, this yields very reasonable agreement with the experimental data without any adjustable parameter. □

Received 23 April; accepted 19 July 2001.

1. Larkin, A. I. Effect of inhomogeneities on the structure of the mixed state of superconductors. *Sov. Phys. JETP* **31**, 784–786 (1970).
2. Fisher, D. S., Fisher, M. P. A. & Huse, D. A. Thermal fluctuations, quenched disorder, phase transitions, and transport in type-II superconductors. *Phys. Rev. B* **43**, 130–159 (1990).
3. Giamarchi, T. & Le Doussal, P. Elastic theory of flux lattices in the presence of weak disorder. *Phys. Rev. B* **52**, 1242–1270 (1995).
4. Zeldov, E. *et al.* Thermodynamic observation of first order vortex lattice melting transition. *Nature* **375**, 373–376 (1995).
5. Deligiannis, K. *et al.* New features in the vortex phase diagram of YBa₂Cu₃O_{7- δ} . *Phys. Rev. Lett.* **79**, 2121–2124 (1997).
6. Jourdain, I. *et al.* Small angle neutron scattering and magnetization measurements in the cubic

(K,Ba)BiO₃ superconductor. *Phys. Rev. Lett.* **82**, 4930–4933 (1999).

prediction (for $R_a < \xi$) for which I is expected to be proportional to $R_a(T)/\lambda^4(T)$. The order–disorder transition temperature can be obtained at 0.4 T as I tends towards zero at $T_m \approx 20$ K (arrow).

7. Paltiel, Y. *et al.* Instabilities and disorder-driven first-order transition of the vortex lattice. *Phys. Rev. Lett.* **85**, 3712–3715 (2000).
8. Marchevsky, M., Higgins, M. J. & Bhattacharya, S. Two coexisting vortex phases in the peak effect regime in a superconductor. *Nature* **409**, 591–594 (2001).
9. Crawford, G. P. & Zumer, S. (eds) *Liquid Crystals in Complex Geometries* (Taylor & Francis, London, 1996).
10. Saunders, K., Radzihovsky, L. & Toner, J. A discotic disguised as a smectic: a hybrid columnar Bragg glass. *Phys. Rev. Lett.* **85**, 4309–4312 (2000).
11. Perruchot, F. *et al.* Hall effect of pinned and depinned 2-D electron and hole solids. *Physica B* **256**, 587–590 (1998).
12. Li, C. C. *et al.* Microwave resonance and weak pinning in two-dimensional hole systems at high magnetic fields. *Phys. Rev. B* **61**, 10905–10909 (2000).
13. Carpentier, D., Le Doussal, P. & Giamarchi, T. Stability of the Bragg glass phase in a layered geometry. *Europhys. Lett.* **35**, 379–384 (1996).
14. Kierfeld, J., Nattermann, T. & Hwa, T. Topological order in the vortex glass phase of high temperature superconductor. *Phys. Rev. B* **55**, 626–629 (1997).
15. Fisher, D. S. Stability of elastic glass phases in random field XY magnets and vortex lattices in type II superconductors. *Phys. Rev. Lett.* **78**, 1964–1967 (1997).
16. Gingras, M. J. P. & Huse, D. A. Topological defects in the random field XY model and the pinned vortex lattice to vortex glass transition in type II superconductors. *Phys. Rev. B* **53**, 15193–15200 (1996).
17. van Otterlo, A., Scalettar, R. T. & Zimanyi, G. T. Phase diagram of disordered vortices from London Langevin simulations. *Phys. Rev. Lett.* **81**, 1497–1500 (1998).
18. Ertas, D. & Nelson, D. R. Irreversibility, mechanical entanglement and thermal melting in superconducting vortex crystals with point impurities. *Physica C* **272**, 79–86 (1996).
19. Giamarchi, T. & Le Doussal, P. Phase diagrams of flux lattices with disorder. *Phys. Rev. B* **55**, 6577–6583 (1997).
20. Fuchs, D. T. *et al.* Transport properties of Bi₂Sr₂CaCu₂O₈ crystals with and without surface barriers. *Phys. Rev. Lett.* **81**, 3944–3974 (1998).
21. Nattermann, T. Scaling approach to pinning: Charge density waves and giant flux creep in superconductors. *Phys. Rev. Lett.* **64**, 2454–2457 (1990).
22. Kim, P. *et al.* Structure of flux line lattices with weak disorder at large length scales. *Phys. Rev. B* **60**, R12589–R12592 (1999).
23. Pardo, F. *et al.* Observation of smectic and moving-Bragg-glass phases in flowing vortex lattices. *Nature* **396**, 348–350 (1998).
24. Troyanovskii, A. M., Aarts, J. & Kes, P. H. Collective and plastic vortex motion in superconductors at high flux densities. *Nature* **399**, 665–668 (1999).
25. Cubitt, R. *et al.* Direct observation of magnetic flux lattice melting and decomposition in the high T_c superconductor BiSrCaCuO. *Nature* **365**, 407–411 (1993).
26. Forgan, E. M. *et al.* Observation by neutron diffraction of the magnetic flux lattice in single-crystal YBaCuO. *Nature* **343**, 735–736 (1990).
27. Yaron, U. *et al.* Structural evidence for a two step process in the depinning of the superconducting flux line lattice. *Nature* **376**, 753–755 (1995).
28. Gammel, P. L. *et al.* Structure and correlations of the flux line lattice in crystalline Nb through the peak effect. *Phys. Rev. Lett.* **80**, 833–836 (1998).
29. Brandt, E. H. Precision Ginzburg–Landau solution of ideal vortex lattices for any induction and symmetry. *Phys. Rev. Lett.* **78**, 2208–2211 (1997).
30. Christen, D. K., Tasset, F., Spooner, S. & Mook, H. A. Study of the intermediate mixed state of niobium by small angle neutron scattering. *Phys. Rev.* **15**, 4506–4509 (1977).

Acknowledgements

We thank C. Simon for discussions.

Correspondence and requests for materials should be addressed to T.K. (e-mail: klein@polycnrs-gre.fr).

The jamming route to the glass state in weakly perturbed granular media

G. D'Anna & G. Gremaud

Institut de Génie Atomique, Département de Physique, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

It has been suggested that a common conceptual framework known as 'jamming' (refs 1 and 2) may be used to classify a wide variety of physical systems; these include granular media³, colloidal suspensions⁴ and glass-forming liquids⁵, all of which display a critical slowdown in their dynamics before a sudden transition to an amorphous rigid state. Decreasing the relevant control parameter (such as temperature, drive or inverse density) may cause geometrical constraints to build up progressively and thus restrict the accessible part of the system's phase space. In glass-forming liquids (thermal molecular systems), jamming is provided by the classical vitrification process of supercooling, characterized by a rapidly increasing and apparently diverging viscosity at sufficiently low temperatures^{6,7}. In driven (athermal) macroscopic systems, a similar slowdown has been predicted to occur, notably in sheared foam or vibrated granular media^{8,9}. Here we report experimental evidence for dynamic behaviour, qualitatively analogous to supercooling, in a driven granular system of macroscopic millimetre-size particles. The granular medium is perturbed by isolated tapping or continuous vibration, with the perturbation intensity serving as a control parameter. We observe the random deflection of an immersed torsion oscillator that moves each time the grains rearrange, like a 'thermometer' sensing the granular noise^{10,11}. We caution that our granular analogy to supercooling is based on similarities in the dynamical behaviour, rather than quantitative theory.

Our experimental method provides evidence for critical slowdown and jamming: first, the noise spectrum is a $1/f^2$ (where f is the frequency) diffusive noise for a perturbation intensity $\Gamma > \Gamma_0$, where $\Gamma_0 \ll \Gamma_f$ is the perturbation intensity at which the granular inverse diffusivity apparently diverges, and Γ_f is the fluidization limit. Second, the divergence is accounted for by an expression like the Vogel–Fulcher–Tammann (VFT) form observed to describe vitrification of glass-forming liquids^{5–7}.

Our granular medium is composed of glass beads of diameter $d_g = 1.1 \pm 0.05$ mm. The granular material is contained in a metallic bucket of 150 mm height and 94 mm diameter, filled to a height of 130 mm. In the experiment, the granular system is subject to a continuous vibration using a vertical shaker driven by a sine wave at the frequency f_s , with f_s selected from the range 50 Hz to 371 Hz. Alternatively, the system is subjected to taps, separated by a time t_w , where a single tap consists of one complete cycle of a f_s sine wave (see lower inset of Fig. 1). An accelerometer fixed on the container measures the intensity of vibration, Γ , as the peak acceleration of the container, normalized to the acceleration of gravity, g . The ideal 'fluidization limit', $\Gamma = 1$, is where single particles start to fly, and we vary Γ in the range $0.003 < \Gamma < 6$. The measured quantity is related to the amplitude a_s and pulsation $\omega_s = 2\pi f_s$ of the sine wave by $\Gamma = a_s \omega_s^2 / g$. To measure the granular noise, we use a torsion oscillator (see upper inset of Fig. 1) whose rotating probe is immersed at a depth L in the granular material. The oscillator performs a brownian-like motion, and the angular deflection, θ , is optically detected. The oscillator is placed on an anti-seismic table and is mechanically isolated from the container, except for the immersed probe. During continuous vibrations, a spectrum analyser captures and processes time-sequences of the angular deflection, $\theta(t)$. In the tapping mode a simple voltmeter is used to measure the angular deflection, θ_n , about t_w seconds after the

end of the tap, with n the number of taps. We note that the grain size is large enough to discard interstitial gas effects¹² and moisture-induced effects^{13,14}. We have taken care always to conduct experiments in well compacted samples (that is, the density is on the 'reversible', steady state branch of the density versus Γ curve in Fig. 2 of ref. 15). Also, the effective experimental timescale is long enough for the granular system to have plenty of time to proceed to the steady state, and ageing effects are discarded.

In the tapping experiment, each deflection point θ_n is measured when all motion induced by the n th tap has stopped. Therefore, each point represents a different static configuration of the granular medium, in which the oscillator is blocked. The power spectrum of a sequence of such deflection points can be considered to be the configuration noise of the perturbed system. Typical tapping power spectra, that is, the squared amplitude of the Fourier transform, denoted $|\theta(f)|^2$, are shown in Fig. 1 for various intensities, Γ . (In Fig. 1, the unit of time is defined as $t = n/113$ s, and f is in Hz, which permits the confrontation with continuous vibration spectra, as shown below.) The noise in Fig. 1 is characterized by $1/f^2$ spectra. Deviations from the $1/f^2$ dependence are observed below about 10 mHz, as the restoring torque of the suspension wires hinders further large deflection of the oscillator. Small deviation also appears at large $\Gamma > 1$, but here we are focusing on the weak Γ regime. A $1/f^2$ noise indicates a diffusive process. Therefore, Fig. 1 shows that the oscillator performs a diffusive, random walk in the granular medium, while the medium itself explores different static

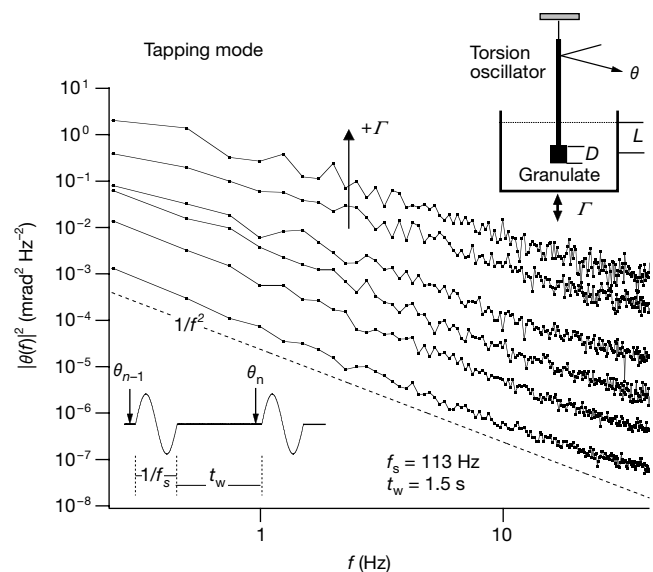


Figure 1 Typical tapping power spectra. Lower inset, a granular medium of glass beads is subject to isolated taps. Upper inset, an immersed torsion oscillator, is used to measure the granular noise: after each tap, when all motion has stopped, the angular deflection, θ , of the oscillator is measured. The power spectra, denoted $|\theta(f)|^2$, of such sequences of deflection points, for various intensities, Γ , of the taps are shown. Each curve is an average of about 10 sequences, 454 points long. The unit of time is defined as the point number times $1/113$ s (see text for details). The values of the intensity of the tap, Γ , are (from the top down) 3.6, 0.93, 0.24, 0.086, 0.50 and 0.025. The other parameters are: $L = 65$ mm, $D = 20.2$ mm, $f_s = 113$ Hz, $t_w = 1.5$ s. Lower inset, each tap is a complete cycle of a sine wave of frequency f_s , and the time separation between taps is denoted t_w . Upper inset, sketch of the torsion oscillator immersed at a depth L into the granular medium. The oscillator has a natural frequency of 20 Hz to 25 Hz, depending on the suspension wires used. The design of the oscillator suspension system permits torsion motion, while other modes are strongly reduced. The probe has an enlarged cylindrical part, of length D , which is covered by a single layer of glass beads, glued on by an epoxy, and the effective radius is 2.5 mm. The spectra of the main panel are the same whether $D = 10.2$ mm or $D = 6$ mm is used.

configurations, driven by the external tapping. The $1/f^2$ noise, and thus the random walk process, is seen for $\Gamma \ll 1$ in Fig. 1, which is an early indication that a weakly perturbed granular medium can display a diffusive behaviour well below the fluidization limit. To

go further we need to determine the exact Γ dependence of the $1/f^2$ noise.

To obtain this information we have used continuous vibrations, which permit averaging over a large number of samples. In fact, the low frequency noise is the same whether continuous vibration or tapping is used, provided that the oscillator is deeply immersed in the granular medium, at say $L > L_c$, where L_c is a critical immersion depth. This is seen in Fig. 2 and Fig. 3. In Fig. 2b, we obtain the same spectrum by a continuous vibration at $f_s = 113$ Hz, or by tapping of the same amplitude, with tap duration $1/113$ s and tap separation $t_w = 1$ s or $t_w = 10$ s. The explanation is that a deeply immersed oscillator experiences a large resisting force from the granular material, because the pressure increases with depth. In such a case, the oscillator cannot move unless the grains rearrange, that is, only during the tap, and the time separation, t_w , of the taps plays no role. (For our purposes, it is not important to determine L_c exactly, since we can always choose $L > L_c$. A detailed study of the L -dependence will be presented elsewhere. We note that, considering the dimension of the container, L always remains in the region where the pressure is linearly dependent on the depth.) The low-frequency noise is the same for all Γ , not only for $\Gamma < 1$ as in Fig. 2b. This can be seen by plotting the low-frequency noise level, denoted $|\theta(1)|^2$, versus Γ , as in Fig. 3, for both continuous vibration and tapping. (Fig. 3 will be discussed in more detail below.) The low-frequency noise using continuous vibrations is the same as in the tapping mode, where deflection points are collected when all motion has ceased and the oscillator is blocked in a given granular configuration; we can therefore decompose the continuous vibration spectra in Fig. 2a into the sum of a configuration noise

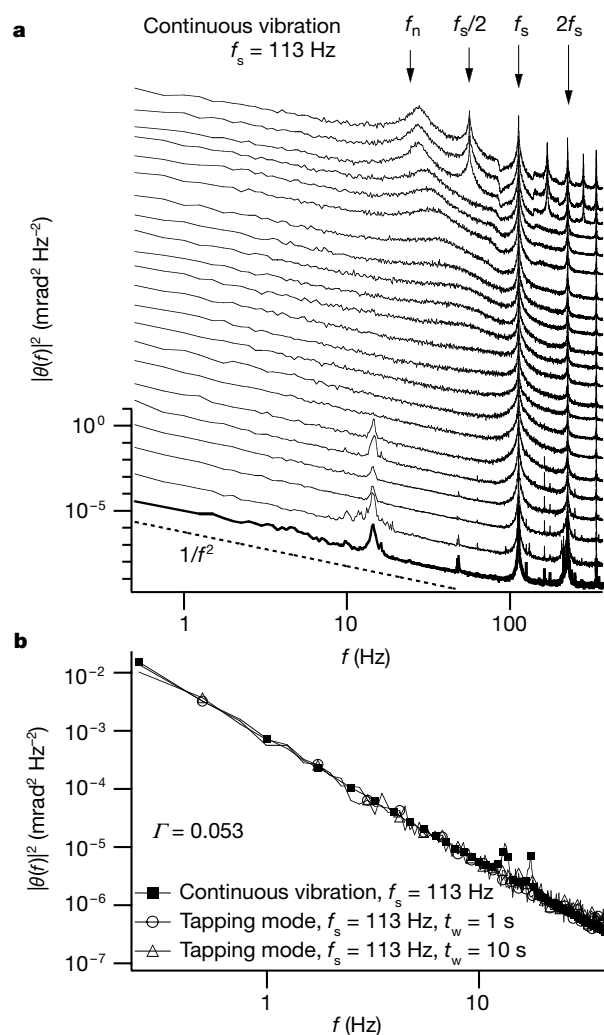


Figure 2 Continuous vibration power spectra and confrontation to the tapping power spectra. **a**, Continuous vibration power spectra, denoted $|\theta(f)|^2$, for various intensities, Γ , of a continuous 113-Hz vibration. Each curve is an average of 100 time-sequences, 4 s long. Except the first bold curve, the curves are shifted vertically for clarity. The intensity Γ is, from the top down, 4.8, 4.0, 3.2, 1.9, 1.2, 0.88, 0.71, 0.61, 0.52, 0.44, 0.38, 0.27, 0.2, 0.14, 0.1, 0.079, 0.059, 0.044, 0.034, 0.026, 0.02 and 0.016. The low-frequency section of the continuous vibration power spectra, below the natural frequency of the oscillator, f_n , is the same as the tapping power spectra measured at the corresponding Γ value (see text for details). The high-frequency noise behaviour, which is inaccessible in the tapping spectra, is characterized by various peaks. First, by increasing Γ a peak grows progressively up and tends to the natural frequency f_n of the oscillator: this is a typical viscous-like feature related to the inertia of the oscillator. Second, peaks at $f_s = 113$ Hz and higher harmonics of f_s are ascribed to the nonlinear coupling of the oscillator to the external vibration through the granular medium. Third, a peak at the subharmonic of f_s at very large Γ is due to a period-doubling effect. **b**, Power spectra at $\Gamma = 0.053$, obtained by continuous vibration and by tapping with different time-separation between the taps. (For clarity, only one symbol every five data values is marked.) The continuous vibration spectrum is obtained with $f_s = 113$ Hz and $\Gamma = 0.053$. The other two are the result of tapping, the tap being of the same amplitude as the continuous vibration and tap duration $1/113$ s. The time separation between the taps is, in one case, $t_w = 1$ s. In the other case, $t_w = 10$ s. The tapping power spectra, with units of time defined as $n/113$ s, overlap with the continuous vibration power spectrum.

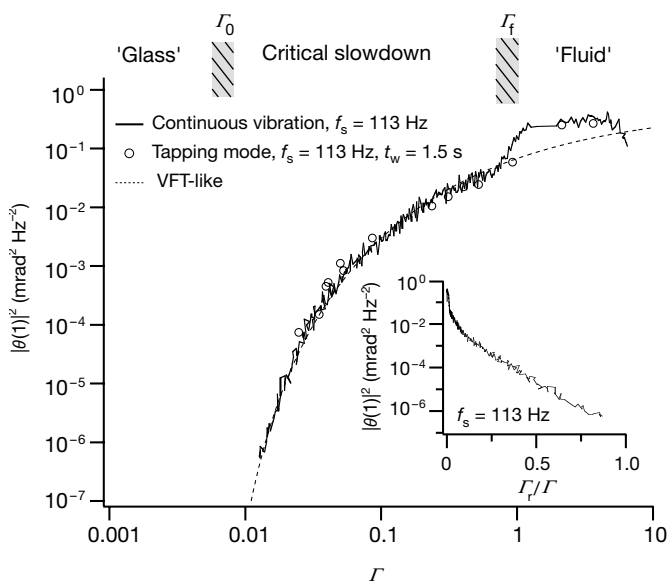


Figure 3 The 'jamming route'. The power spectrum level at 1 Hz, denoted $|\theta(1)|^2$, versus Γ , obtained from continuous vibration measurements, similar to those of Fig. 2a. The continuous vibration curve is composed of 290 points, one point collected each 144 seconds. These data are obtained by decreasing Γ , but no difference is seen if Γ is increased. Also, no difference is seen if the points are collected at a rate two orders of magnitude smaller. Some points (open symbols) are obtained from tapping spectra, as in Fig. 1. The sharp change in the noise level at Γ , apparently in the range 0.7 to 1.1, indicates the fluidization threshold. The dotted line is obtained according to a VFT-like form (see text for details). Γ_0 is the perturbation intensity where the $1/f^2$ noise level, and thus the configuration diffusivity, extrapolates to zero. Inset, the same data as in the main panel in a semilogarithmic plot of $|\theta(1)|^2$ versus Γ_r/Γ , where Γ_r is a reference vibration intensity at which $|\theta(1)|^2$ reaches the value 4×10^{-7} mrad² Hz⁻². These data clearly show the analogy between the granular behaviour and the physics of glass-forming liquids that supercool.

component, which dominates at low frequency, plus a high-frequency noise component. The high-frequency noise behaviour is, of course, inaccessible in the tapping spectra, because it is determined by the motion of the massive oscillator during the taps. These fast dynamics give rise to viscous-like features in the spectra (see also Fig. 2 legend), but, although it is of interest, we shall not discuss the high frequency noise further here.

We now focus on the low-frequency noise and address the Γ dependence. Figure 3 illustrates the main findings. First, a feature at Γ_f , in the range 0.7 to 1.1, is apparently the signature of the vibration-induced fluidization. The origin of this feature, which is not seen in usual glass-forming systems, is unknown at the moment, but it might be a specific feature of the so-called fluidization transition, or be related to a partial crystallization induced by the surfaces of the container. Second, the configuration noise apparently approaches zero critically, that is, the inverse noise level diverges. Both data points obtained by continuous vibration and tapping are shown in Fig. 3. The critical approach to zero can be described by a modified VFT form $|\theta(1)|^2 = A \exp[B(\Gamma - \Gamma_0)^p]$, with $A = 0.5$, $B = -2.0$, $\Gamma_0 = 0.005$ and $p = -0.4$, where Γ_0 is the vibration intensity at which $|\theta(1)|^2$ extrapolates to zero. The same data are also shown in the inset of Fig. 3 in a semilogarithmic plot of $|\theta(1)|^2$ versus Γ_r/Γ , where Γ_r is a reference vibration intensity at which $|\theta(1)|^2$ reaches a fixed, small value. Accordingly, for $\Gamma \ll 1$ an Arrhenius behaviour also fits the data. A VFT-like divergence is suggested by the complete curve below Γ_f . We note that, by the Wiener–Khinchine theorem, for a $1/f^2$ noise, the value of the noise at a fixed frequency, for example, $|\theta(1)|^2$, is proportional to the diffusion coefficient (see, for example, ref. 16). Therefore, Fig. 3 establishes a VFT-like diffusive slowdown. If we view Fig. 3 as a ‘phase diagram’, the emerging picture is that, by decreasing the drive below the fluidization limit, Γ_f , the perturbed granular medium evolves into a final amorphous, rigid, ‘glass’ state at Γ_0 in much the same way as when the viscosity diverges, or as when the diffusivity approaches zero, for supercooled liquids^{5–7} in general. This ‘jamming route’ is expected to be the common signature of a wide variety of other thermal and athermal systems, where the fluctuation-accessible phase space is progressively reduced^{1,2}.

The Γ -controlled diffusive slowdown behaviour reported here, and the analogy with the physics of supercooled liquids, might provide some insight into the possible effective temperature in vibrated granular systems^{10,11}. The observed exponent, $p = -0.4$, which is close to -0.5 , suggests that an effective temperature should scale as $\Gamma^{0.5}$, as then the divergence could be described by a standard VFT form, that is, with $p = -1$. Indeed, previous susceptibility measurements¹⁷ have suggested the role of a control parameter $\tau \propto \Gamma^{0.5}$. However, the question remains open, in our opinion, since an operational definition of the temperature must, at least, be constructed on an effective fluctuation-dissipation theorem¹⁸.

Models based on ideal hard spheres have been widely used to simulate liquids and glasses¹⁹. A thermal system can be seen as subjected to a random (Langevin) force, with a white spectrum¹⁶. In the granular medium, the random force includes all sorts of vibration modes generated by a single tap, including those induced by the falling grains themselves, and the elastic vibration of the container. Therefore, we expect a ‘coloured’ random-force spectrum (that is, with a broad maximum at some characteristic frequency) strongly dependent on the tap duration $1/f_s$ and the tap amplitude a_s , but also dependent on experimental parameters such as the container form and material, the total weight of the granular material, the granular elastic and contact (tribological) properties. After each tap, these vibrations determine the details of the glass configuration in which the grains will be trapped; thus they determine the ‘freezing’ limit Γ_0 , which characterizes the effective ‘jamming route’. The parameters selected here are not critical, in the sense that a similar critical slowdown (but with a different value of Γ_0) is observed if, for example, f_s or the total granular weight are changed. □

Received 17 May; accepted 15 August 2001.

1. Liu, A. J. & Nagel, S. R. Jamming is not just cool anymore. *Nature* **396**, 21–22 (1998).
2. Liu, A. J. & Nagel, S. R. (eds) *Jamming and Rheology: Constrained Dynamics on Microscopic and Macroscopic Scales* (Taylor and Francis, New York, 2001).
3. Jaeger, H. M., Nagel, S. R. & Behringer, R. P. Granular solids, liquids, and gases. *Rev. Mod. Phys.* **68**, 1259–1273 (1996).
4. Trappe, V., Prasad, V., Cipelletti, L., Segre, P. N. & Weitz, D. A. Jamming phase diagram for attractive particles. *Nature* **411**, 772–775 (2001).
5. Tarjus, G. & Kivelson, D. in *Jamming and Rheology: Constrained Dynamics on Microscopic and Macroscopic Scales* (eds Liu, A. J. & Nagel, S. R.) 20–38 (Taylor and Francis, New York, 2001).
6. Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
7. Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).
8. Langer, S. A. & Liu, A. J. Sheared foam as a supercooled liquid? *Europhys. Lett.* **49**, 68–74 (2000).
9. O’Hern, C. S., Langer, S. A., Liu, A. J. & Nagel, S. R. Force distributions near jamming and glass transitions. *Phys. Rev. Lett.* **86**, 111–114 (2001).
10. Kurchan, J. Emergence of macroscopic temperatures in systems that are not thermodynamical microscopically: towards a thermodynamical description of slow granular rheology. *J. Phys. Condens. Matter* **12**, 6611–6617 (2000).
11. Barrat, A., Kurchan, J., Loreto, V. & Sellitto, M. Edwards’ measures for powders and glasses. *Phys. Rev. Lett.* **85**, 5034–5037 (2000).
12. Pak, H. K., Van Doorn, E. & Behringer, R. P. Effects of ambient gases on granular materials under vertical vibrations. *Phys. Rev. Lett.* **74**, 4643–4646 (1995).
13. Bocquet, L., Charlaix, E., Ciliberto, S. & Crassous, J. Moisture-induced ageing in granular media and the kinetics of capillary condensation. *Nature* **396**, 735–737 (1998).
14. D’Anna, G. Dissipative dynamics of granular media investigated by forced torsion pendulum. *Europhys. Lett.* **51**, 293–299 (2000).
15. Nowak, E. R., Knight, J. B., Ben-Naim, E., Jaeger, H. M. & Nagel, S. R. Density fluctuations in vibrated granular materials. *Phys. Rev. E* **57**, 1971–1982 (1998).
16. Risken, H. *The Fokker-Planck Equation* (Springer, Berlin, 1989).
17. D’Anna, G. & Gremaud, G. Activated-like hopping transition in weakly vibrated granular media. *Europhys. Lett.* **54**, 599–604 (2001).
18. Berthier, L., Cugliandolo, L. F. & Iguain, J. L. Glassy systems under time-dependent driving forces: Application to slow granular rheology. *Phys. Rev. E* **63**, 051302–1–051302–15 (2001).
19. Speedy, R. J. The hard sphere glass transition. *Mol. Phys.* **95**, 169–178 (1998).

Acknowledgements

We thank P. Martin, A. Barrat, L. Cugliandolo and J. Kurchan for discussions and comments.

Correspondence and requests for materials should be addressed to G.D.
(e-mail: danna@igahpse.epfl.ch).

Photochemical cycling of iron in the surface ocean mediated by microbial iron(III)-binding ligands

K. Barbeau*†, E. L. Rue‡, K. W. Bruland‡ & A. Butler*

* Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA

‡ Institute of Marine Sciences, University of California, Santa Cruz, California 95064, USA

Iron is a limiting nutrient for primary production in large areas of the oceans^{1–4}. Dissolved iron(III) in the upper oceans occurs almost entirely in the form of complexes with strong organic ligands^{5–7} presumed to be of biological origin^{8,9}. Although the importance of organic ligands to aquatic iron cycling is becoming clear, the mechanism by which they are involved in this process remains uncertain. Here we report observations of photochemical reactions involving Fe(III) bound to siderophores—high-affinity iron(III) ligands produced by bacteria to facilitate iron acquisition^{10–12}. We show that photolysis of Fe(III)–siderophore complexes leads to

† Present address: Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

the formation of lower-affinity Fe(III) ligands and the reduction of Fe(III), increasing the availability of siderophore-bound iron for uptake by planktonic assemblages. These photochemical reactions are mediated by the α -hydroxy acid moiety, a group which has generally been found to be present in the marine siderophores that have been characterized^{13–15}. We suggest that Fe(III)-binding ligands can enhance the photolytic production of reactive iron species in the euphotic zone and so influence iron availability in aquatic systems.

Siderophores or their breakdown products probably make up a large component of the strong Fe(III)-binding ligands that dominate iron(III) speciation in surface ocean waters. Numerous marine cyanobacteria and heterotrophic bacteria have been shown to produce siderophores in culture in response to iron stress^{10–15}. Voltammetric studies have shown that the ferric-ion binding

strength of siderophores is similar to that of the naturally occurring strong Fe(III)-binding ligands in sea water^{6,16}. Most recently, Fe(III)-binding ligands isolated from the California coastal upwelling system were shown to possess functional groups characteristic of siderophores¹⁷. We suggest that bacteria associated with particles or other diffusively limited microenvironments in the water column may produce and utilize siderophores as an iron-uptake mechanism, and that microbially colonized particles could then serve as a source of siderophores to the bulk aqueous phase via diffusion or disaggregation¹⁴.

In light of the evidence for siderophores contributing to the classes of Fe(III)-binding ligands present in sea water, we used structurally characterized marine siderophores, the aquachelins, as model compounds to probe the upper-ocean biogeochemistry of iron, specifically addressing the potential photoreactivity of

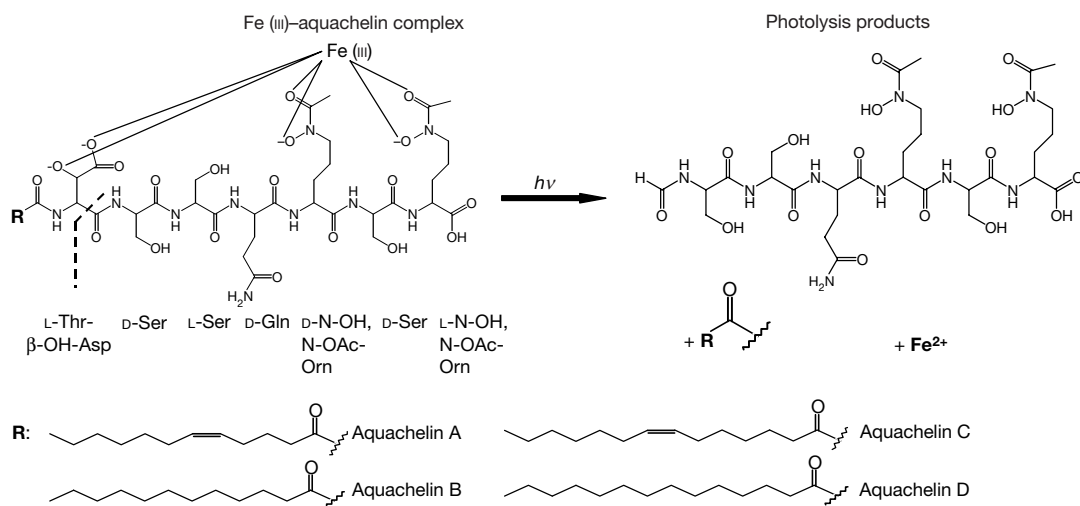


Figure 1 Schematic representation of the photochemical reaction of Fe(III)-aquachelin complexes. The dashed line indicates the position of photolytic cleavage of the aquachelin ligand. Iron(III) is reduced to iron(II) via ligand-to-metal charge transfer. Orn, ornithine.

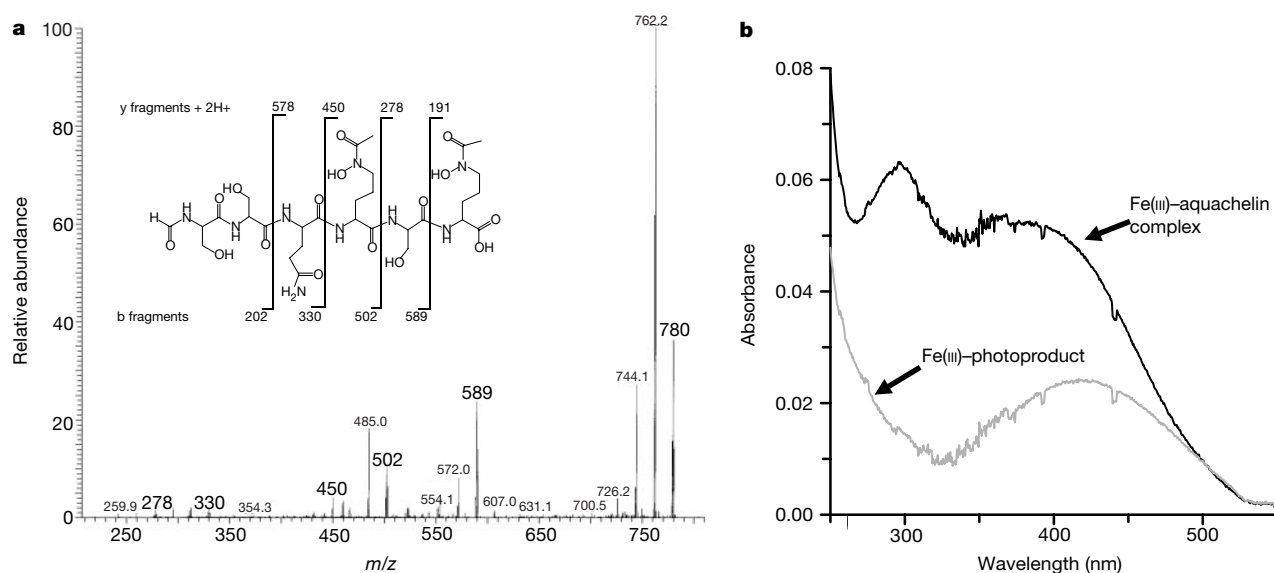


Figure 2 Tandem mass spectrometry and UV-visible spectrophotometry of the Fe(III)-aquachelin peptide photoproduct. **a**, Tandem mass spectrometry of the Fe(III)-aquachelin C peptide photoproduct (m/z 780) confirms loss of β -hydroxyaspartate. “y” and “b” nomenclature refers to charge retained by the carboxy-terminal fragment or the amino-terminal fragment of the peptide, respectively. **b**, UV-visible spectrophotometry of Fe(III)-

aquachelin C and of the Fe(III)-photoproduct indicates loss of an electronic transition centred at 300 nm; this corresponds to the charge-transfer band from β -hydroxyaspartate to iron(III). The hydroxamate-Fe(III) charge-transfer band is still present, indicating that Fe(III)-binding groups remain on the peptide photoproduct.

naturally occurring, strong Fe(III)–ligand complexes. The aquachelins are a suite of siderophores produced by the heterotrophic marine bacterium *Halomonas aquamarina* strain DS40M3, which was isolated from a water sample taken from 40 m depth (within the euphotic zone, mixed layer) over the continental slope in the eastern equatorial Atlantic Ocean^{14,15}. The aquachelins each contain one of a series of fatty-acid tails and a peptidic headgroup that forms a 1:1 Fe(III):ligand complex via two hydroxamate groups and one β -hydroxyaspartate residue¹⁵ (Fig. 1). The aquachelins are among the siderophores produced by open-ocean marine bacteria that have been fully structurally characterized. We note that most of these siderophores (that is, the alterobactins, aerobactin, marinobactins, aquachelins, and others; refs 13–15, and K.B. and A.B., unpublished results) contain the α -hydroxy acid group that mediates the ligand-to-metal charge-transfer reaction reported here.

Fe(III)–aquachelin complexes readily undergo photolysis in natural sunlight, resulting in partial ligand breakdown and changes in iron speciation (Fig. 1). A similar reaction pathway exists for the iron(III) complexes of the marinobactins, a suite of marine siderophores structurally related to the aquachelins, with fatty-acid tails and the same Fe(III)-binding groups, but a different headgroup structure¹⁵. In contrast to their photochemical reactivity, the ferric ion complexes of the aquachelins and marinobactins are structurally stable on timescales of weeks to months in sterile sea water in the dark, even at elevated temperatures (50 °C).

Photolysis of Fe(III)–aquachelin C results in oxidative cleavage of the ligand at the site of the β -hydroxyaspartate residue (Fig. 1). Two primary ligand photoproducts are formed, one more hydrophilic (the peptide portion) and one more hydrophobic (the fatty-acid component) than the original Fe(III)–aquachelin complex, as indicated by reverse-phase high-performance liquid chromatography (RP-HPLC). Electrospray ionization mass spectrometry (ESI-MS) of the hydrophilic photoproduct revealed a molecular ion peak (mass to charge ratio $m/z = 780$) corresponding to the mass of the peptide headgroup minus the fatty-acid tail and the β -hydroxyaspartate moiety. As expected on the basis of their molecular structures, the other aquachelins in the series (A, B, D) formed the same hydrophilic peptide photoproduct (m/z 780) upon sunlight irradiation of their ferric-ion complexes.

We confirmed the absence of the β -hydroxyaspartate residue from the peptide photoproduct by amino-acid analysis and by tandem mass spectrometry (Fig. 2a). The peptide photoproduct

retains the two hydroxamate groups, and is capable of binding iron(III), as shown by the detection of an iron adduct (m/z 833) in ESI-MS. We performed measurements of the Fe(III)-binding conditional stability constant ($K_{\text{FeL,Fe}'}^{\text{cond}}$) of the Fe(III)–aquachelin peptide photoproduct via competitive ligand equilibration-adsorptive cathodic stripping voltammetry (CLE-ACSV)⁶. These measurements indicated a significant drop in Fe(III)-binding strength relative to the original siderophore; $K_{\text{FeL,Fe}'}^{\text{cond}}$ was $10^{12.2} \text{ M}^{-1}$ for aquachelin B, and was $10^{11.5} \text{ M}^{-1}$ for the peptide photoproduct of aquachelin B. This decrease in Fe(III)-binding strength is consistent with partial loss of the ligating ability of the siderophore ligand, due to loss of the β -hydroxyaspartate moiety as determined by structural analysis of the peptide photoproduct. The value of $K_{\text{FeL,Fe}'}^{\text{cond}}$ of the intact siderophore correlates with field observations of the strong Fe(III)-binding ligand L_1 in sea water (average $K_{\text{FeL,Fe}'}^{\text{cond}} = 10^{12.5 \pm 0.3} \text{ M}^{-1}$), while the $K_{\text{FeL,Fe}'}^{\text{cond}}$ of the photoproduct is similar to values observed for the weaker class of Fe(III)-binding ligands in sea water, L_2 (average $K_{\text{FeL,Fe}'}^{\text{cond}} = 10^{11.6 \pm 0.2} \text{ M}^{-1}$)^{6,18}.

The photoreaction of Fe(III)–aquachelin complexes is initiated by a β -hydroxyaspartate to iron(III) ligand-to-metal charge-transfer reaction. The ultraviolet–visible spectrum of photolysed Fe(III)–aquachelin C (Fig. 2b) demonstrates loss of an electronic transition in the near-ultraviolet (wavelength, 300 nm), which probably corresponds to the charge transfer from β -hydroxyaspartate (an α -hydroxy acid moiety) to iron(III). The photoreactivity of iron(III)

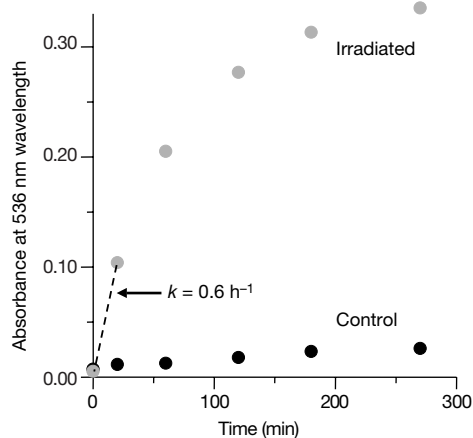


Figure 3 Rate of iron(II) production during photolysis of Fe(III)–aquachelin C in natural sunlight, as determined via absorption of the ferrous tris(BPDS) complex at 536 nm. Significant concentrations of the ferrous tris(BPDS) complex were formed only in the irradiated solution (grey dots), and not in the dark control (black dots). The initial rate of ferrous tris(BPDS) complex formation in the irradiated solution was 0.6 h^{-1} .

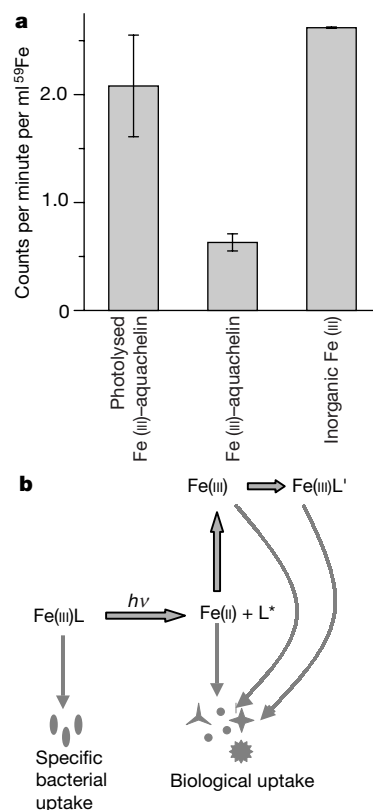


Figure 4 Biological uptake of iron from photolysed Fe(II)–aquachelin complexes, and diagram of siderophore-mediated photochemical iron cycling. **a**, Uptake experiment performed with a natural assemblage of organisms obtained from the oligotrophic Atlantic. ^{59}Fe uptake by organisms $>0.2 \mu\text{m}$ in size is shown from ^{59}Fe –aquachelin B complexes (intact and previously photolysed) and from inorganic ^{59}Fe additions, over a 2-day incubation. **b**, Schematic summary of photochemical iron cycling in sea water as mediated by photoreactive Fe(III)–siderophore complexes, with consequences for biological iron uptake. FeL represents the original Fe(III)–siderophore complex, L^* is the ligand photoproduct, and L' is another iron(III) ligand.

complexes of simple polycarboxylate molecules containing α -hydroxy acids (for example, citrate and oxalate) is well known, leading to ligand oxidation and iron(III) reduction^{19,20}. The present work is, to our knowledge, the first report of similar reactivity of an α -hydroxy acid group in a hexadentate siderophore molecule: this is a significant finding because α -hydroxy acids are present in numerous polydentate siderophores and phytosiderophores from terrestrial and aquatic environments²¹. Most of the pelagic marine siderophores that have been structurally characterized to date contain α -hydroxy acid moieties that participate in iron(III) binding^{13–15}, and when ferrated these siderophores do undergo photolysis as observed for the Fe(III)–aquachelins (K.B. and A.B., unpublished results). Thus photoreactivity may be a common characteristic of oceanic Fe(III)–siderophore complexes. Photolysis of the Fe(III)–aquachelins is mediated by light in the ultraviolet–A range of the solar spectrum. The reaction readily occurs in polycarbonate vessels in natural sunlight, and proceeds even when light of wavelength less than 385 nm is filtered out (determined via irradiation with neutral-density filters, data not shown). These results show that photolysis of Fe(III)– α -hydroxy acid siderophore complexes is mediated by light across the solar ultraviolet spectrum, and therefore these reactions may occur deep into the euphotic zone of open ocean waters²².

As expected on the basis of a ligand-to-metal charge-transfer reaction mechanism, photolysis of ferric aquachelin complexes results in the reduction of iron(III) to iron(II). We trapped iron(II) formed during irradiation of a solution of Fe(III)–aquachelin C in natural sunlight by using the Fe(II)-specific chelating agent bathophenanthroline disulphonate (BPDS; Fig. 3). Approximately 70% of the total iron in solution was trapped by BPDS in sunlight (versus 5% in the dark control), consistent with a photoreductive reaction mechanism. The pseudo-first-order rate constant for Fe(II) production over the first 20 minutes of irradiation was 0.6 h^{-1} , within the range of the fastest rate constants reported for colloidal Fe oxide photoreduction in sea water²³ and more than fast enough to support turnover of Fe(III)–siderophore complexes on a daily timescale. In additional trapping experiments, we added BPDS immediately after photolysis to avoid possible artefacts due to BPDS photosensitization or photoreduction of Fe(III)–BPDS species. Under these conditions (see Methods), after complete photolysis of ferric aquachelin in solution, up to 60% of the total iron was detected as the ferrous tris(BPDS) complex. In light of current knowledge regarding the dominance of organic complexes in open-ocean iron(III) speciation^{5–7,18}, these results (showing the photoproduction of iron(II) from pelagic marine bacterial iron(III)–siderophore complexes containing the α -hydroxy acid group) may serve to explain observations of diel iron(II) (or reactive iron) cycling in oceanic surface waters^{23–25}. This is an advance over earlier attempts to experimentally model Fe photocycling in the open ocean on the basis of colloidal iron oxides^{22,23}, a form of iron now recognized as unlikely to be significant in truly oligotrophic oceanic regimes.

Photolysis of Fe(III)–siderophore complexes results in oxidation of the ligand and reduction of iron(III), thus probably affecting the bioavailability of siderophore-complexed iron(III). To test this hypothesis, we performed an ⁵⁹Fe uptake experiment using a natural assemblage of planktonic organisms from the oligotrophic Atlantic Ocean to compare the relative bioavailability of ⁵⁹Fe from additions of intact ⁵⁹Fe(III)–aquachelin B, previously photolysed ⁵⁹Fe(III)–aquachelin B, and inorganic ⁵⁹Fe(III) (FeCl₃) (see Methods). The ⁵⁹Fe(III) uptake results (Fig. 4a) show that photolysis did increase the bioavailability of ⁵⁹Fe(III) added as the aquachelin B complex. As expected, the intact ⁵⁹Fe(III)–aquachelin B complex was largely unavailable to the natural assemblage. In contrast, uptake of ⁵⁹Fe(III) from previously photolysed ⁵⁹Fe(III)–aquachelin B was indistinguishable from incubations in which inorganic ⁵⁹FeCl₃ was added.

The results reported here show that microbial Fe(III) chelates probably facilitate extensive photochemical cycling of iron in ocean surface waters (Fig. 4b). The key features of this cycle involve photolysis of Fe(III)–ligand complexes, with reduction of iron(III) to iron(II) and oxidation of the ligand ($L \rightarrow L^*$). The fate of photolytically produced iron(II) could include direct biological uptake, oxidation to iron(III), or oxidation with subsequent complexation by another strong iron(III) ligand (L'). (Possible chelation of iron(II) by organic ligands and the resulting effect on iron(II) oxidation rates is an important question for further study²⁶). The photo-oxidized ligand (L^*) still retains Fe(III)-binding capacity and so could continue to contribute to the upper-ocean iron(III) ligand pool, possibly as part of a weaker ligand class^{6,18}. Iron bound by L^* may be more available for biological uptake, as the conditional stability constant of Fe(III) L^* is reduced relative to that of the original ferric ligand (L). The potential effects of photolysis on Fe(III)–ligand complexes in recently upwelled water masses could be another important component of this cycle, operative over longer timescales²⁷.

The cycle described above is likely to be a fundamental feature of the marine biogeochemistry of iron, strongly influencing the interaction of marine biota with this critical micronutrient. Our results indicate that α -hydroxy acid-containing microbial iron(III) chelates in the upper ocean may play a dual role, both participating in siderophore-mediated bacterial iron transport and also enhancing the photolytic production of reactive iron species in the euphotic zone. These findings have implications for the cycling of iron in the oceans. Given the occurrence of α -hydroxyacid-containing hexadentate siderophores in a variety of near-surface soil and aquatic environments²¹, the photoreactivity reported here advances our understanding of how this major class of siderophores functions in biological iron acquisition. □

Methods

Isolation and structural characterization of photoproducts

The isolation and structural characterization of aquachelin siderophores has been described elsewhere¹⁵. Photoproducts were generated by irradiating solutions of ferrated aquachelins in 0.2- μm -filtered, UV-irradiated Sargasso seawater (UVSSW) at pH 8. Concentrations of Fe(III)–aquachelin in solution ranged from 20 to 40 μM , with a 10–15% excess of free ligand. Irradiations were performed in natural sunlight in acid-washed quartz flasks, in a temperature-controlled water bath (18 °C), for a duration of 6–8 h.

Photoproducts were isolated and purified directly from photolysed solutions via RP-HPLC on Vydac C-4 and C-18 columns. ESI-MS of peptide photoproduct isolates was performed using a standard ESI source, with single quadrupole mass analyser. The amino-acid composition of the peptide photoproduct was determined with Marfey's reagent (*N*- α -(2,4-dinitro-5-fluorophenyl)-L-alanine amide) (ref. 28), and by tandem mass spectrometry using both ion trap (Fig. 2a) and triple quadrupole mass spectrometers. UV–visible spectra of intact and photolysed Fe(III)–aquachelin solutions in UVSSW were obtained on a Cary 3e spectrophotometer.

CLE/ACSV analysis

Stock solutions of aquachelin B were prepared by adding 500 μl of CH₃OH (Fisher Optima grade) to the concentrated siderophore sample. Secondary standards were made up in Milli-Q water such that a 10 nM solution of aquachelin B could be made in a 10 ml sample of UVSSW—seawater collected from Monterey Bay from which trace metals and trace-metal-binding organic compounds have been removed²⁹. Increasing iron additions were performed on the aquachelin B samples, and left to equilibrate for 1 h. Iron speciation was measured by competitive ligand equilibration (CLE)—followed by adsorptive cathodic stripping voltammetry (ACSV) with salicylaldoxime (SA) as the competing ligand⁶. This CLE-ACSV technique detects the concentration of the electrochemically active, metal-added ligand complex (Fe(SA)₂) formed after a competing equilibrium has been established between Fe(III)' (where Fe(III)' refers to the sum of all inorganic Fe(III) species such as Fe(OH)₃⁺, Fe(OH)₂⁺ or Fe(OH)₄[−], salicylaldoxime, and aquachelin B. Total dissolved iron was measured by acidifying the sample (pH 1.7), followed by briefly microwaving and, once cooled to room temperature, applying the ACSV technique⁶.

Fe(II) trapping experiments

BPDS trapping of Fe(II) formed during photolysis of Fe(III)–aquachelin complexes was performed by irradiating a solution of BPDS and Fe(III)–aquachelin in natural sunlight for 4.5 h (a solution of 76 μM BPDS, 25 μM Fe(III)–aquachelin C in UVSSW). Samples were taken at t_0 and at regular intervals (20–90 min) during irradiation; the absorbance of the

ferrous tris(BPDS) complex was measured at 536 nm (molar extinction coefficient, $\epsilon = 1.927 \times 10^4 \text{ l mol}^{-1} \text{ cm}^{-1}$). A portion of the BPDS/Fe(III)–aquachelin solution was kept in the dark and served as the control.

Additional trapping experiments were performed by adding BPDS immediately after photolysis. Solutions of Fe(III)–aquachelin C (10 μM) in UVSSW at pH 5 were irradiated (20–30 min) using an artificial light source (300-W Hg lamp with infrared filter and Pyrex UV filter). Immediately after irradiation, BPDS was added to aliquots of the photolysed solution (BPDS final concentration 300 μM). Solutions were allowed to equilibrate for 5–10 min with BPDS, then absorbance was measured at 536 nm. A non-irradiated portion of the Fe(III)–aquachelin solution served as a dark control.

Natural assemblage iron-uptake experiment

^{59}Fe –aquachelin complexes were formed by adding $^{59}\text{FeCl}_3$ in 0.5 M HCl (NEN Life Sciences) to aquachelin B in UVSW (Fe:ligand ratio 1:5; pH circumneutral). This stock was allowed to equilibrate overnight at room temperature. For photolysed ^{59}Fe –aquachelin additions, ^{59}Fe –aquachelin stock solution was diluted in Milli-Q water to a concentration of 250 nM Fe:1,250 nM aquachelin and irradiated in an acid-washed quartz flask under natural sunlight for 8 h. This mixture was then immediately added to incubation bottles (final concentration 5 nM Fe). For non-photolysed ^{59}Fe –aquachelin additions, ^{59}Fe –aquachelin B stock was added for a final concentration of 5 nM Fe, 25 nM aquachelin. For inorganic ^{59}Fe incubations, $^{59}\text{FeCl}_3$ was added directly for a final concentration of 5 nM Fe.

Sea water for incubations was collected from 80 m depth (chlorophyll maximum) over the Bermuda rise in the North Atlantic Ocean. Water was collected in Go-Flo bottles using trace-metal-clean techniques, dispensed into replicate acid-washed 1-litre polycarbonate bottles, and tracers added under a laminar-flow bench. Bottles were incubated in an on-deck flow-through incubator at 1–2% ambient light level. After 2 d, aliquots were removed, filtered onto 0.2- μm polycarbonate filters, and rinsed with Ti(III)–citrate–EDTA reagent to remove extracellular ^{59}Fe (ref. 30). Filters were counted on a Cobra II γ -ray counter to <5% counting error. Values presented are the mean $\pm$ one standard deviation of replicate bottles.

Received 10 January; accepted 30 July 2001.

- Martin, J. H. *et al.* Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. *Nature* **371**, 123–129 (1994).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- Hutchins, D. A. & Bruland, K. W. Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime. *Nature* **393**, 561–564 (1998).
- Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
- Gledhill, M. & Van den Berg, C. M. G. Determination of complexation of iron (III) with natural organic complexing ligands in seawater using cathodic stripping voltammetry. *Mar. Chem.* **47**, 41–54 (1994).
- Rue, E. L. & Bruland, K. W. Complexation of iron(III) by natural organic ligands in the Central North Pacific as determined by a new competitive ligand equilibration/adsorptive cathodic stripping voltammetric method. *Mar. Chem.* **50**, 117–138 (1995).
- Wu, J. & Luther, G. W. Complexation of Fe(III) by natural organic ligands in the Northwest Atlantic Ocean by a competitive ligand method and a kinetic approach. *Mar. Chem.* **50**, 159–177 (1995).
- Hutchins, D. A., Witter, A. E., Butler, A. & Luther, G. W. Competition among marine phytoplankton for different chelated iron species. *Nature* **400**, 858–861 (1999).
- Johnson, K. S., Gordon, R. M. & Coale, K. H. What controls dissolved iron concentrations in the world ocean? *Mar. Chem.* **57**, 137–161 (1997).
- Trick, C. G. Hydroxamate-siderophore production and utilization by marine eubacteria. *Curr. Microbiol.* **18**, 375–378 (1989).
- Wilhelm, S. & Trick, C. G. Iron-limited growth of cyanobacteria: Multiple siderophore production is a common response. *Limnol. Oceanogr.* **39**, 1979–1984 (1994).
- Granger, J. & Price, N. M. The importance of siderophores in the iron nutrition of heterotrophic marine bacteria. *Limnol. Oceanogr.* **44**, 541–555 (1999).
- Reid, R. T., Live, D. H., Faulkner, D. J. & Butler, A. A siderophore from a marine bacterium with an exceptional ferric ion affinity constant. *Nature* **366**, 455–458 (1993).
- Haygood, M. G., Holt, P. D. & Butler, A. Aerobactin production by a planktonic marine *Vibrio* sp. *Limnol. Oceanogr.* **38**, 1091–1097 (1993).
- Martinez, J. S. *et al.* Self-assembling amphiphilic siderophores from marine bacteria. *Science* **287**, 1245–1247 (2000).
- Lewis, B. L. *et al.* Voltammetric estimation of iron(III) thermodynamic stability constants for catecholate siderophores isolated from marine bacteria and cyanobacteria. *Mar. Chem.* **50**, 179–188 (1995).
- Macrellis, H. M., Trick, C. G., Rue, E. L. & Bruland, K. W. Collection and detection of natural iron-binding ligands from seawater. *Mar. Chem.* (in the press).
- Rue, E. L. & Bruland, K. W. The role of organic complexation on ambient iron chemistry in the equatorial Pacific Ocean and the response of a mesoscale iron addition experiment. *Limnol. Oceanogr.* **42**, 901–910 (1997).
- Zuo, Y. & Hoigne, J. Formation of hydrogen peroxide and depletion of oxalic acid in atmospheric water by photolysis of iron(III)–oxalato complexes. *Environ. Sci. Technol.* **26**, 1014–1022 (1992).
- Faust, B. C. & Zepp, R. G. Photochemistry of aqueous iron(III)–polycarboxylate complexes: Roles in the chemistry of atmospheric and surface waters. *Environ. Sci. Technol.* **27**, 2517–2522 (1993).
- Albrecht-Gary, A.-M. & Crumbliss, A. L. In *Metal Ions in Biological Systems* Vol. 35, Iron Transport and Storage in Microorganisms, Plants, and Animals (eds Sigel, A. & Sigel, H.) 239–327 (Marcel Dekker, New York, 1998).
- Wells, M. L., Mayer, L. M., Donard, O. F. X., de Souza Sierra, M. M. & Ackelson, S. G. The photolysis of colloidal iron in the oceans. *Nature* **353**, 248–250 (1991).

- Johnson, K. S., Coale, K. H., Elrod, V. A. & Tindale, N. W. Iron photochemistry in seawater from the equatorial Pacific. *Mar. Chem.* **46**, 319–334 (1994).
- O'Sullivan, D. W., Hanson, A. K., Miller, W. L. & Kester, D. R. Measurement of Fe(II) in surface water of the equatorial Pacific. *Limnol. Oceanogr.* **36**, 1727–1741 (1991).
- Waite, T. D., Szymczak, R., Espey, Q. I. & Furnas, M. J. Diel variations in iron speciation in northern Australian shelf waters. *Mar. Chem.* **50**, 79–91 (1995).
- Emmenegger, L., King, D. W., Sigg, L. & Sulzberger, B. Oxidation kinetics of Fe(II) in a eutrophic Swiss lake. *Environ. Sci. Technol.* **32**, 2990–2996 (1998).
- Archer, D. E. & Johnson, K. A model of the iron cycle in the ocean. *Glob. Biogeochem. Cycles* **14**, 269–279 (2000).
- Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg. Res. Commun.* **49**, 591–569 (1984).
- Donat, J. R. & Bruland, K. W. Direct determination of dissolved cobalt and nickel in seawater by differential pulse cathodic stripping voltammetry preceded by adsorptive collection of cyclohexane-1,2-dione dioxime complexes. *Anal. Chem.* **60**, 240–244 (1988).
- Hudson, R. J. M. & Morel, F. M. M. Distinguishing between extra- and intracellular iron in marine phytoplankton. *Limnol. Oceanogr.* **34**, 1113–1120 (1989).

Acknowledgements

This work was supported by the NSF/DOE Environmental Molecular Science Institute (A.B. and K.W.B.), the National Institutes of Health (A.B.) and the University of California President's postdoctoral fellowship program (K.B.).

Correspondence and requests for materials should be addressed to A.B. (e-mail: butler@chem.ucsb.edu).

Earliest presence of humans in northeast Asia

R. X. Zhu*, K. A. Hoffman†, R. Potts‡, C. L. Deng*, Y. X. Pan*, B. Guo*, C. D. Shi*, Z. T. Guo*, B. Y. Yuan*, Y. M. Hou§ & W. W. Huang§

* Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100101, China

† Physics Department, California Polytechnic State University, San Luis Obispo, California 93410, USA

‡ Human Origins Program, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA

§ Institute of Vertebrate Palaeontology and Palaeoanthropology, Chinese Academy of Sciences, Beijing 100044, China

The timing of the earliest habitation and oldest stone technologies in different regions of the world remains a contentious topic in the study of human evolution^{1–4}. Here we contribute to this debate with detailed magnetostratigraphic results on two exposed parallel sections of lacustrine sediments at Xiaochangliang in the Nihewan Basin, north China; these results place stringent controls on the age of Palaeolithic stone artifacts that were originally reported over two decades ago⁵. Our palaeomagnetic findings indicate that the artifact layer resides in a reverse polarity magnetozone bounded by the Olduvai and Jaramillo subchrons. Coupled with an estimated rate of sedimentation, these findings constrain the layer's age to roughly 1.36 million years ago. This result represents the age of the oldest known stone assemblage comprising recognizable types of Palaeolithic tool in east Asia, and the earliest definite occupation in this region as far north as 40° N.

The chronology of east Asian Palaeolithic tools is pertinent to both the oldest human occupation in this region and the overall framework of human origin and migration in the Old World. The age of Palaeolithic sites in east Asia has been controversial because of the absence of suitable material for accurate isotopic dating. However, Cenozoic continental deposits have provided abundant mammalian fossils, especially in China, and offer a solid base for biostratigraphic and magnetostratigraphic study. The Nihewan

Basin (formerly Nihowan), one in a series of east Asian Cenozoic basins, is an excellent example (Fig. 1).

The basin, which is located in the transition zone between the North China Plain and the Inner Mongolian Plateau, is filled with Pliocene to Holocene lacustrine and wind-blown deposits. Covering an area of roughly 150–200 km², it provides extensive sedimentary exposures that are more than 90 m thick. Since the 1920s, the basin has attracted the attention of geologists and palaeontologists because of its well-developed late Cenozoic strata that are rich in mammalian fossils, known as the Nihewan Fauna. These fossils traditionally correspond to the Villafranchian fauna in Europe^{6,7}.

Several early Pleistocene sites reported from this basin are relevant to understanding early occupation in east Asia. These sites, from the area near the eastern margin of the basin, include Xiaochangliang (XCL; 40.2° N, 114.65° E)⁵ and Donggutuo (DGT)⁸, and have been claimed to represent some of the earliest occupations of *Homo erectus* in eastern Asia⁹. Stone artifacts excavated from the XCL archaeological site^{5,10,11} are made primarily of fine and coarse-grained chert and volcanic rock, and rarely of quartz, quartzite and sandstone, which are all obtained from local outcrops.

The studied assemblage ($n = 3,113$) consists of flakes and flaked pieces, including clearly identifiable tool or core forms that are recognizable as standard classes of Palaeolithic flaked artifacts^{5,10–12}.

These include mainly side scrapers and notches and a few examples of end scrapers, burins and disc cores (Fig. 2). Stone flaking and retouch were done by direct free-hand and, rarely, by bipolar percussion. Because the assemblage consists predominantly of unmodified flakes and flake fragments, and cores, retouched flakes and formal tool classes are relatively rare, some analyses^{9,10,13} have emphasized the informality and simple character of the XCL industry. Others^{11,14–16}, however, have noted that the presence of nearly cylindrical cores, blades and flat core platforms (suggestive of platform preparation) indicate an advanced state beyond the oldest known stone technologies.

Comprising greyish-white clayey silts to silty sands with a thickness of 0.5–0.8 m, the XCL stone artifact layer is located along cliffs of the southern bank of the Sanggan River and lies about 66.5 m below the Holocene soil (Figs 1 and 3). The layer preserves a rich vertebrate fauna, including *Allophaiomys* cf. *A. pliocaenicus*, *Mimomys chinensis*, *Hyaena* (*Pachycrocuta*) *licenti*, *Palaeoloxodon* sp., *Hipparion* sp., *Proboscideiparian sinensis*, *Equus sanmeniensis*, *Coelodonta antiquitatis*, *Martes* sp., *Cervus* sp. and *Gazella* sp.^{5,17}. Of these Nihewan faunal elements, the first six taxa are indicative of a late Pliocene to early Pleistocene age^{5,17–20}. *E. sanmeniensis* ranges from early to middle Pleistocene^{17,18}, whereas other taxa continued to the end of the Pleistocene (for example, *C. antiquitatis*), or are

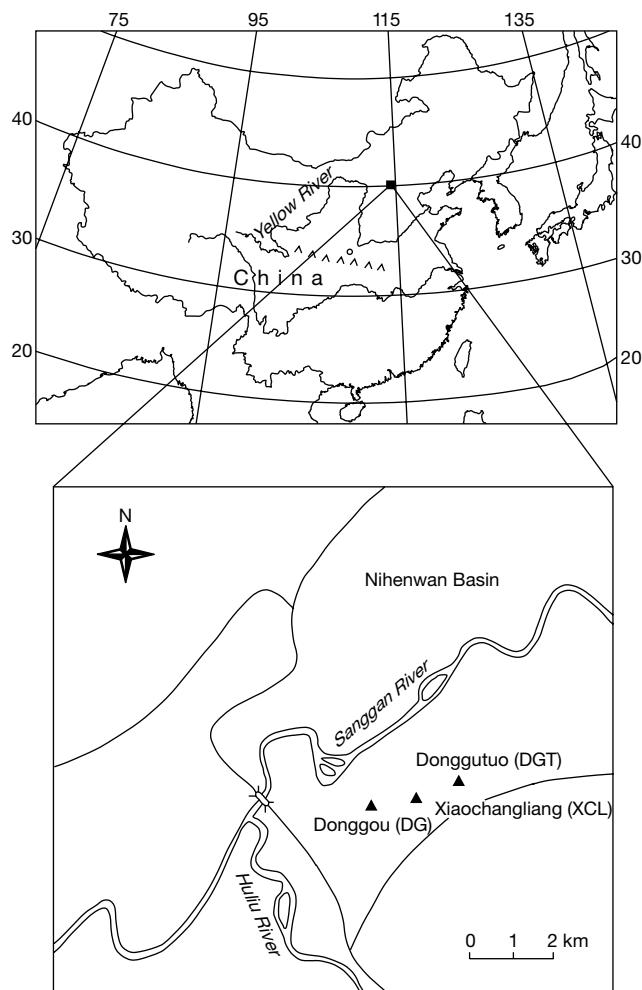


Figure 1 Location of the Xiaochangliang Palaeolithic site in the Nihewan Basin. Two other sites mentioned in the text, Donggutuo and Donggou, are indicated. North China lies north of the Qinling Mountains (▲▲), which correspond to the southern limit of the Loess Plateau. The Lantian site (open circle), mentioned in the text, is found on the southern Loess Plateau along the Yellow River, which is the principal river system south and west of the Nihewan Basin.

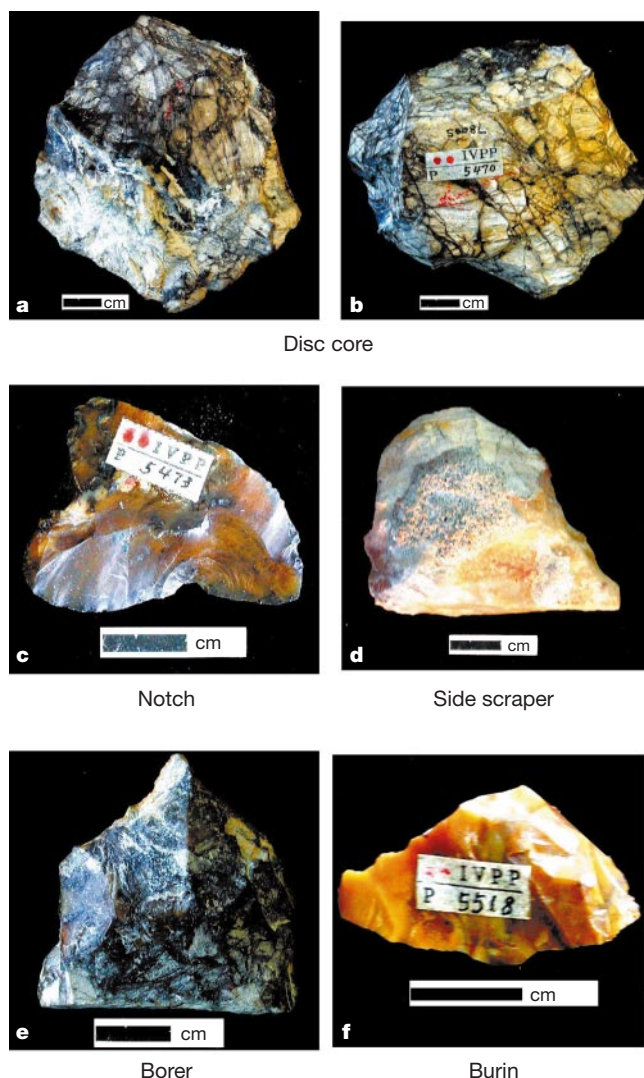


Figure 2 Examples of flaked stone artifacts excavated from Xiaochangliang. **a, b**, Opposite faces of a disc core; **c**, notch; **d**, side scraper; **e**, borer; **f**, burin.

insufficiently identified (for example, *Marctes* sp.) to provide a useful age constraint. Clearly, a precise age for the XCL site cannot be determined on the basis of fauna alone.

We examined the 73-m-thick XCL section palaeomagnetically. A 90-m-thick parallel section named Donggou (DG), 1 km away from the XCL section, was studied for comparison (see Fig. 1). Three parallel specimens with independent orientation were sampled

from each horizontal level in the two sections. Sampling intervals were 25 and 35 cm, which provided 280 and 257 horizon levels at the XCL and DG sections, respectively. Close proximity of the two sections facilitated precise stratigraphic correlation and effectively minimized uncertainty resulting from lateral facies variation. The two sections have an almost identical sequence of sedimentary facies containing pronounced sedimentological marker layers (Fig. 3).

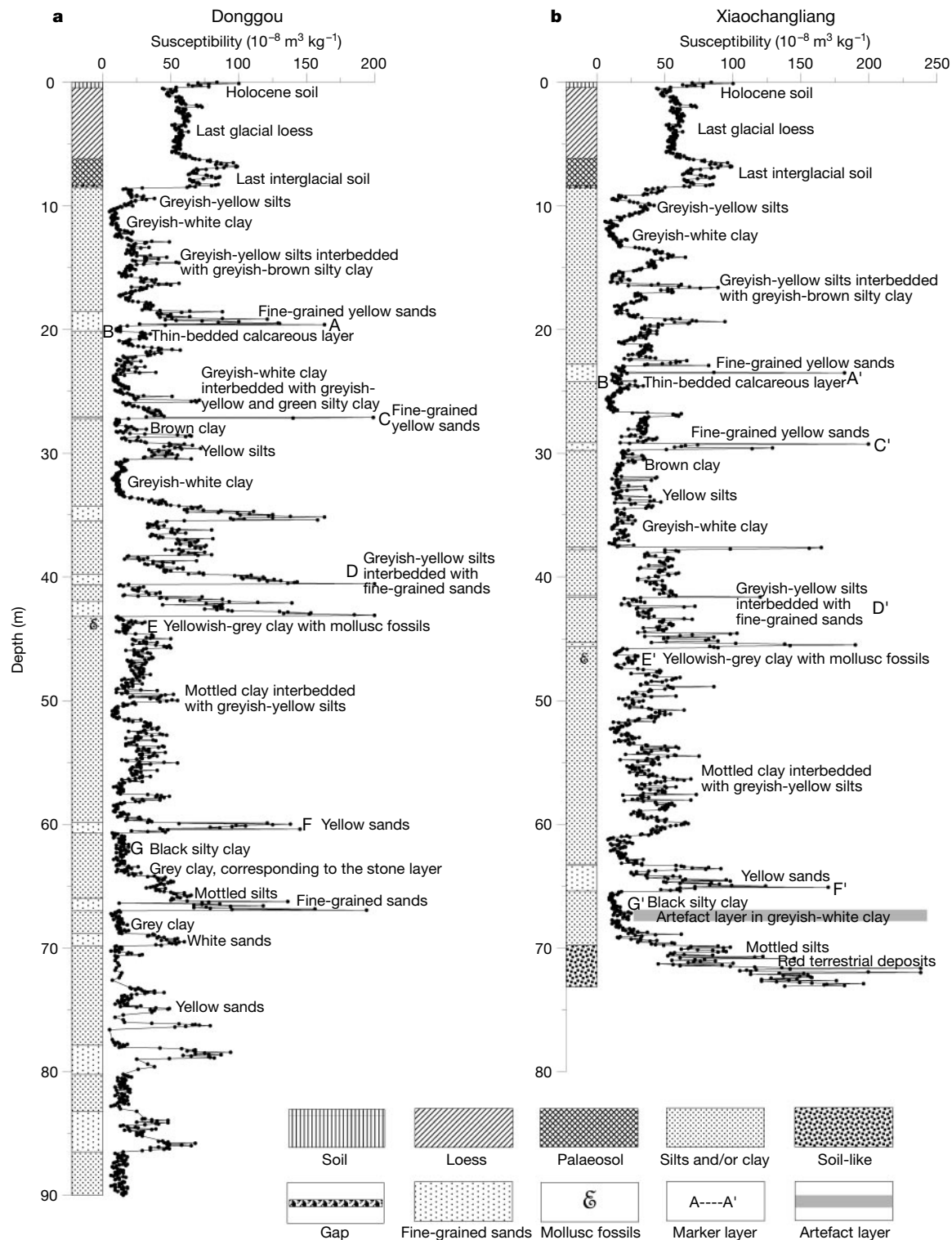


Figure 3 Lithologic and magnetic susceptibility stratigraphy of Donggou (a) and Xiaochangliang (b) sections. The two sections have an almost identical sequence of sedimentary facies containing pronounced sedimentological marker layers, A–G.

Notably, a distinctive black silty clay layer (2 cm thick) found directly above the artifact layer is also found in the DG section at a depth of 62.20 m. In addition, several fine-grained sand layers, associated with high values of magnetic susceptibility, are found at the same position within the polarity framework (Figs 3 and 4). Therefore, the XCL section can be correlated with the DG section on lithological and stratigraphic grounds.

Before attempting to establish the polarity stratigraphy by demagnetization of the natural remanent magnetization (NRM), we checked the reliability of the two palaeomagnetic records using various rock magnetic methods, including anisotropy of magnetic susceptibility. The results were positive, indicating that pure magnetite of pseudo-single domain size is the principal carrier of the magnetic remanence and that the sedimentary magnetic fabric had been unperturbed since deposition.

The averaged characteristic remanent magnetization vector direction for each set of three parallel specimens after thermal and/or alternating field demagnetization procedures yielded virtual geomagnetic pole latitudes, along with error estimates, that were subsequently used to define the succession of magnetostratigraphic polarity in the two sections (Fig. 4). At the DG section there are five magnetozones: three normal, N1 (0–18.72 m), N2 (45.20–48.77 m) and N3 (81.12 m to the bottom); and two reversed, R1 (18.72–45.20 m) and R2 (48.77–81.12 m). At the XCL section there are four magnetozones: two normal, N1 (0–20.95 m) and N2 (46.55–49.35 m); and two reversed, R1 (20.95–46.55 m) and R2 (49.35 m to the bottom). Hence, apart from the fact that the DG section seems to extend further back in time, thereby exposing the N3 normal magnetozones, the magnetostratigraphic and lithologic findings indicate that the two sections had experienced very similar

depositional histories. The only possible sedimentation gap found was in N1 (Fig. 3), between the top of the lacustrine sediments and the bottom of the aeolian deposits at a depth of 8.5 m. Note that the layer containing Palaeolithic stone artifacts at XCL is found 17.6 m below the bottom of magnetozone N2 (Fig. 4).

Both sections of lacustrine sediment seem to be virtually continuous. Indeed, in comparison with other types of continental deposit, lake sediment such as that at XCL is more likely to accumulate continuously and less likely to be perturbed. Given that the Holocene soil, Malan loess, and soil associated with the last interglacial are found to overlay the lacustrine sections, the magnetozones determined for XCL and DG can be correlated to the geomagnetic polarity timescale²¹. Specifically, magnetozones N1, N2 and N3 correspond to the Brunhes chron, the Jaramillo subchron and the Olduvai subchron, respectively—a correlation that is further strengthened by the fact that the mammalian fauna indicates a late Pliocene to early Pleistocene age^{17–20}.

Although the onset of magnetozone R2 at XCL was inaccessible, the entire magnetozone was successfully sampled at DG where its thickness is 32.35 m. As the duration of magnetozone R2 (that is, the time interval between the termination of the Olduvai subchron and the onset of the Jaramillo subchron) is known to span roughly 0.7 Myr²¹, the average rate of accumulation of this section turns out to be about 4.6 cm kyr⁻¹. Hence, given that the 2-cm layer of black silty clay in section DG (layer G in Fig. 3), which correlates with the top of the artifact layer in XCL, lies 13.4 m below the Jaramillo onset, we calculate the age of that layer to be about 1.36 Myr.

This result documents the oldest archaeological record in the Nihewan Basin and provides a new perspective regarding early human populations in east Asia. Although earlier stone tools may be

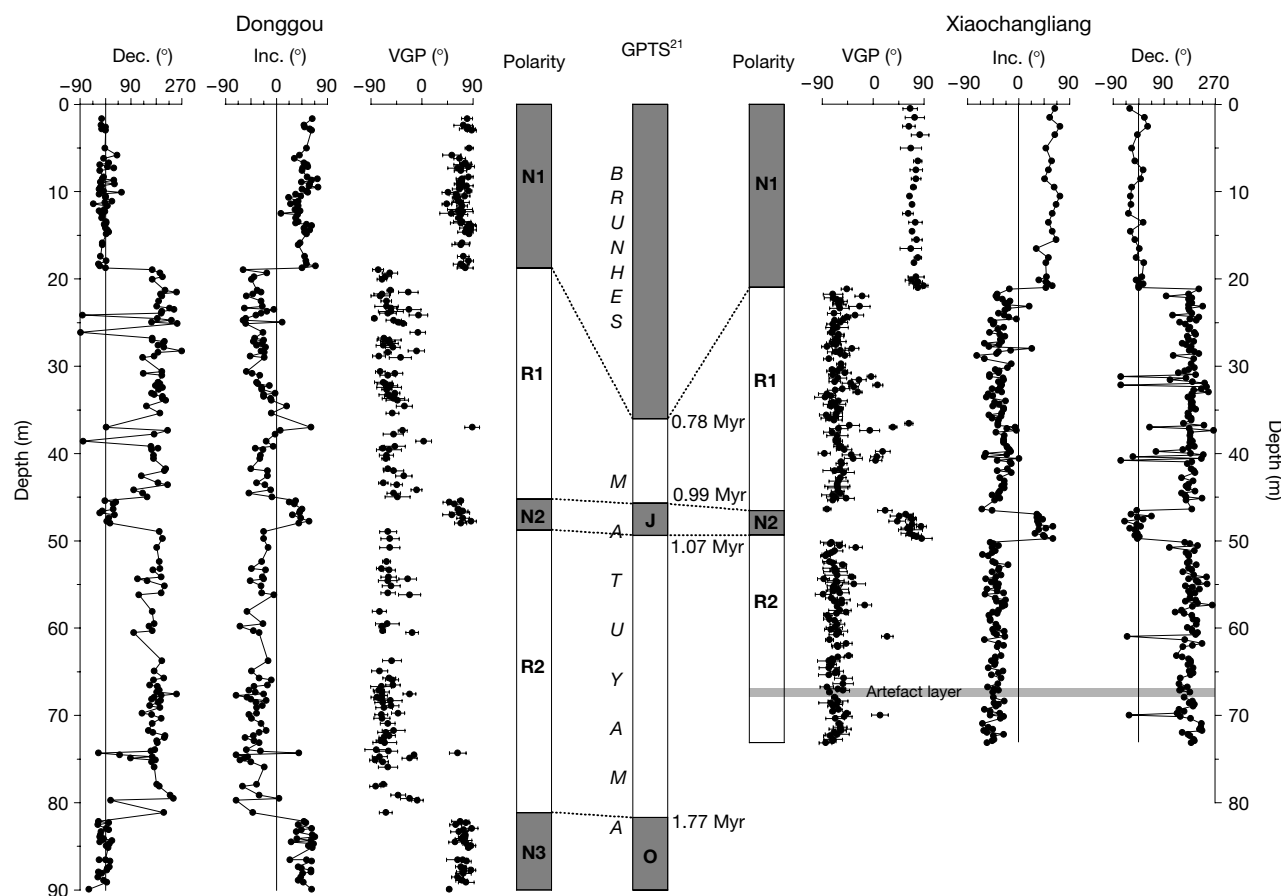


Figure 4 Determined magnetostratigraphy at Donggou (DG) and Xiaochangliang (XCL) and correlation with the geomagnetic polarity timescale²¹. Inc., inclination; Dec., declination; VGP, virtual geomagnetic pole.

present in southern China²², our age estimate provides evidence for the oldest known artifact assemblage consisting of indisputable tool forms in east Asia. It also denotes the oldest unambiguous presence of early humans in east Asia at a latitude of at least 40°N, following the range extension to this same latitude in western Eurasia by about 1.7 Myr³. The spread of toolmakers to this latitude implies that early Pleistocene human populations in east Asia were able to adapt to diverse climatic settings. It also suggests that populations might have dispersed to and from east Asia over a broad latitudinal range, although highlands of the Qinghai–Tibetan Plateau probably limited a northern migration route. The age of the XCL Palaeolithic site is, furthermore, about 0.2 Myr earlier than *Homo erectus* (about 1.15 Myr) found at Lantian²³, located on the middle Yellow River, roughly 900 km southwest of the Nihewan Basin (Fig. 1). These two localities suggest that populations were able to occupy or shift their range over a considerable area, from Nihewan to the southern margin of the Loess Plateau, during a time of enhanced global and regional climatic variability that included intermittent aridification of north China^{24,25}. □

Methods

Rock magnetic methods and results

Stepwise acquisition of isothermal remanent magnetization was investigated using an alternating gradient field magnetometer up to 2.2 T and pulse magnetometer up to 2.7 T. For 95% of the samples studied, 90% of the saturation magnetization was acquired below 0.3 T, showing that a low-coercivity magnetic carrier dominates the remanence. After removing the coarser fraction (> 100 µm), we used thermomagnetic analysis of the bulk sediment to determine a Curie point of 580 °C, which is characteristic of pure magnetite. The saturation magnetization (M_s), saturation remanence (M_{rs}), coercivity field (H_c) and the coercivity of the remanence (H_{cr}) were determined and their ratios combined in a Day plot²⁶. This plot indicates that the magnetite grain sizes are predominantly pseudo-single domain. Details of the magnetic properties of the carriers will be addressed elsewhere.

Anisotropy of magnetic susceptibility

Anisotropy of magnetic susceptibility (AMS) measurements were performed using a Kappabridge KLY-3s (Geofyzika Brno). The susceptibility tensor for each sample was calculated from measurements in 15 positions, by a described method²⁷. To avoid potential problems associated with heating²⁸, we completed the AMS measurements before any thermal demagnetization was conducted. In the 666 samples studied, the magnetic lineation (L) was found to be smaller than the magnetic foliation (F), indicating that the AMS ellipsoid is oblate. Most minimum susceptibility axes (K_{min}) are close to the vertical, perpendicular to the bedding plane ($I_{mean} = 78.3^\circ$), and tightly grouped, whereas the inclinations of the maximum axes (K_{max}) are very shallow ($I_{mean} = 6.7^\circ$). These results are typical for an original sedimentary magnetic fabric that has been unperturbed since deposition.

Demagnetization of the NRM

Remanence measurements were made using a 2G three-axis cryogenic magnetometer installed in field-free space (< 300 nT). We performed stepwise alternating field and thermal demagnetization techniques on 1,611 samples taken from the two sections. Both methods could isolate the characteristic remanent magnetization after removing one or two soft components of magnetization (see Supplementary Information). Total 223 (80%) and 207 (81%) sampling levels with 669 and 621 oriented samples gave reliable characteristic remanence directions at XCL and DG, respectively. NRM intensities for 669 samples at XCL range from 30.6 to 1,792 mA m⁻¹ with a mean value of 285.2 mA m⁻¹; similarly, for 621 samples at DG the intensities range from 29.6 to 1,770 mA m⁻¹ with a mean value of 248.9 mA m⁻¹.

Received 4 June; accepted 26 July 2001.

- Morwood, M. J. O. *et al.* Fission-track ages of stone tools on the east Indonesian island of Flores. *Nature* **392**, 173–176 (1998).
- Hou, Y. M. *et al.* Mid-Pleistocene Acheulean-like stone technology of the Bose basin, South China. *Science* **287**, 1622–1626 (2000).
- Gabunia, L. *et al.* Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: Taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
- Ambrose, S. H. Paleolithic technology and human evolution. *Science* **291**, 1748–1753 (2001).
- You, Y. Z., Tang, Y. J. & Li, Y. Discovery of the Palaeoliths from the Nihewan Formation. *Quat. Sinica* **5**, 1–11 (1980).
- Barbour, G. B. The deposits of the Sankanh Valley. *Bull. Geol. Soc. China* **4**, 53–55 (1925).
- Teilhard de Chardin, P. & Piveteau, J. Les mammifères fossiles de Nihowan (Chine). *Ann. Paléontol.* **19**, 1–154 (1930).
- Wei, Q. New Paleolithic site from the Nihewan beds. *Acta Anthropol. Sinica* **4**, 223–231 (1985).
- Schick, K. D. *et al.* Archaeological perspectives in the Nihewan Basin, China. *J. Hum. Evol.* **21**, 13–26 (1991).

- Chen, C. *et al.* 1998 excavation of the Xiaochangliang site at Yang-yuan, Hebei. *Acta Anthropol. Sinica* **18**, 225–239 (1999).
- Keates, S. G. *Early and Middle Pleistocene Hominid Behaviour in Northern China* BAR International Series Vol. 863, 107 (John and Hedges, Oxford, 2000).
- Chen, C. *et al.* Lithic analysis of the Xiaochangliang industry. *Acta Anthropol. Sinica* (in the press).
- Li, Y. On the progressiveness of the stone artifacts from the Xiaochangliang site at Yuanguan, Hebei. *Acta Anthropol. Sinica* **18**, 241–254 (1999).
- Pei, W. Z. Comment on 'Discovery of the Palaeoliths from the Nihewan Formation'. *Quat. Sinica* **5**, 11–12 (1980).
- Jia, L. P. Comment on 'Discovery of the Palaeoliths from the Nihewan Formation'. *Quat. Sinica* **5**, 12 (1980).
- Huang, W. W. On the stone industry of Xiaochangliang. *Acta Anthropol. Sinica* **4**, 301–307 (1985).
- Tang, Y. J., Li, Y. & Chen, W. Y. Mammalian fossil and the age of Xiaochangliang paleolithic site of Yuanguan, Hebei. *Vertebrata Palasiatica* **33**, 74–83 (1995).
- Wei, Q. in *Evidence for Evolution—Essays in Honor of Prof. Chungchien Yong on the Hundredth Anniversary of His Birth* (eds Tong, Y. *et al.*) 193–207 (China Ocean, Beijing, 1997).
- Huang, W. P., Fan, Q. R. in *Wushan Hominid Site 105–109* (Ocean, Beijing, 1991).
- Qiu, Z. X. Nihewan fauna and Q/N boundary in China. *Quat. Res.* **20**, 142–154 (2000).
- Berggren, W. A. *et al.* A revised Cenozoic geochronology and chronostratigraphy in time scales and global stratigraphic correlations: A unified temporal framework for an historical geology. in *Geochronology, Timescales, and Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V., Aubry, M. & Hardenbol, J.) *SEPM spec. publ.* **54**, 129–212 (1995).
- Huang, W. P. *et al.* Early Homo and associated artifacts from Asia. *Nature* **378**, 275–278 (1995).
- An, Z. S. & Ho, C. K. New magnetostratigraphic dates of Lantian *Homo erectus*. *Quat. Res.* **32**, 213–221 (1989).
- Liu, T. S. & Ding, Z. L. Comparison of Plio-Pleistocene climatic changes in different monsoonal regions and implications for human evolution. *Quat. Sci.* **4**, 289–298 (1999).
- Rutherford, S. & D'Hondt, S. Early onset and tropical forcing of 100,000-year Pleistocene glacial cycles. *Nature* **408**, 72–75 (2000).
- Day, R., Fuller, M. & Schmidt, V. A. Hysteresis properties of titanomagnetites: grain size and compositional dependence. *Phys. Earth. Planet. Inter.* **13**, 260–267 (1977).
- Jelinek, V. Statistical processing of anisotropy of magnetic susceptibility measured on groups of specimens. *Studia Geoph. Et Geod.* **22**, 50–62 (1978).
- Rochette, P., Jackson, M. M. & Aubourg, C. Rock magnetism and the interpretation of anisotropy of magnetic susceptibility. *Rev. Geophys.* **30**, 209–226 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the National Natural Science Foundation of China.

Correspondence and requests for materials should be addressed to R.X.Z. (e-mail: rxzhu@mail.c-geos.ac.cn) or K.A.H. (e-mail: khoffman@calpoly.edu).

Lagged effects of ocean climate change on fulmar population dynamics

Paul M. Thompson* & Janet C. Ollason††

* University of Aberdeen, Department of Zoology, Lighthouse Field Station, Cromarty, Ross-shire IV11 8YJ, UK

† University of Aberdeen, Department of Zoology, Culterty Field Station, Newburgh, Ellon, Aberdeenshire AB41 6AA, UK

Environmental variation reflected by the North Atlantic Oscillation affects breeding and survival in terrestrial vertebrates^{1,2}, and climate change is predicted to have an impact on population dynamics by influencing food quality or availability³. The North Atlantic Oscillation also affects the abundance of marine fish and zooplankton^{4,5}, but it is unclear whether this filters up trophic levels to long-lived marine top predators. Here we show by analysis of data from a 50-year study of the fulmar that two different indices of ocean climate variation may have lagged effects on population dynamics in this procellariiform seabird.

† Present address: Nether Ardrain, Ellon, Aberdeenshire AB41 8RA, UK.

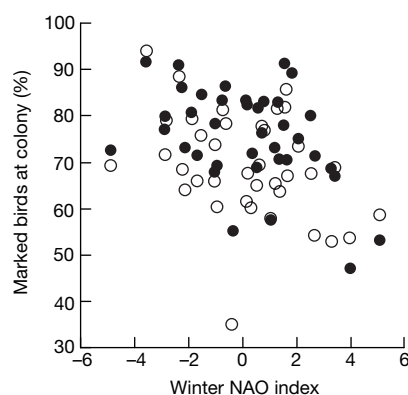


Figure 1 Annual variation in the percentage of colour-ringed adult fulmars that were observed at the colony during each breeding season in relation to the annual North Atlantic Oscillation index for that summer. Data are presented separately for males (filled

circles) and females (open circles). All adults used in the analysis were known to have bred and were later confirmed to be alive in that season.

Annual variability in breeding performance is influenced by the North Atlantic Oscillation, whereas cohort differences in recruitment are related to temperature changes in the summer growing season in the year of birth. Because fulmars exhibit delayed reproduction, there is a 5-year lag in the population's response to these effects of environmental change. These data show how interactions between different climatic factors result in complex dynamics, and that the effects of climate change may take many years to become apparent in long-lived marine top predators.

The northern fulmar, *Fulmarus glacialis*, is one of the most abundant seabirds in the North Atlantic and has provided a classic example of species range expansion in the past two centuries⁶. Although some think that this expansion may have resulted from increased discards from whaling and fishing activities⁷, others have argued that the pattern of expansion is better explained by oceanographic changes^{8,9}. Uncertainty over the foraging areas used by birds from different breeding colonies has prevented explicit tests of these hypotheses^{9,10}. Furthermore, environmental variability is most likely to affect reproductive success^{11,12}, or the survival of juveniles rather than adults¹³. Because these long-lived seabirds do not recruit to breeding colonies until they are several years old¹⁴, a time series of several decades may be required to assess the demographic consequences of the factors influencing those early life stages.

Since 1950, studies of fulmar population ecology have been made on Eynhallow, Scotland¹⁵, offering a rare opportunity to explore the effects of climate change on seabird colony dynamics. The proportion of breeding adults present at the colony each summer was related significantly to the winter North Atlantic Oscillation (NAO) index (Fig. 1) for both males (analysis of variance (ANOVA) $F_{1,36} = 8.30$, $P < 0.01$, $r^2 = 0.19$) and females ($F_{1,36} = 4.17$, $P < 0.05$, $r^2 = 0.10$), with the lower probability of recording females at the colony reflecting sex differences in foraging cycles¹⁴. The absence of breeding adults in any one year could result from unrecorded early nest failures or intermittent breeding^{16,17}, either of which could be related to the NAO through its impact on secondary production in the northeast Atlantic⁴. Alternatively, variations in wind speed markedly affect the cost of flight in these soaring birds¹⁸, and the occurrence of weaker westerlies during a negative NAO¹⁹ may simply increase the probability of birds being observed at the colony. Whatever the ultimate cause, the NAO's influence on the attendance patterns of breeding adults clearly needs to be considered when making inferences about the size of a population from a census of nesting pairs.

More significantly for their demography, hatching success and fledging success in the following summer were both related

significantly to the winter NAO index (hatching success, $F_{1,34} = 5.23$, $P < 0.05$, $r^2 = 0.14$, Fig. 2a; fledging success, $F_{1,36} = 6.52$, $P < 0.05$, $r^2 = 0.15$, Fig. 2b). Multivariate analyses incorporating environmental factors, density and levels of disturbance indicated that colony size did not influence these reproductive parameters; however, the winter NAO and previously reported effects of disturbance together explained almost 30% of the observed variation in annual fledging success ($F_{2,32} = 6.92$, $P < 0.01$, $r^2 = 0.28$). The abundance of overwintering *Calanus* in the North Sea increases during negative NAO conditions⁴, and this may provide a link between these observed effects on reproductive success through an increase in availability of the crustaceans and fish that are known to be taken by fulmars in Scottish waters²⁰. Alternatively, poor reproductive success may reflect the severe winter conditions experienced in low NAO years¹⁹.

Variation in the proportion of fledged chicks that subsequently recruited to the colony was investigated for cohorts born between 1958 (when long-lasting monel rings were first used) and 1980 (to allow sufficient time to detect new breeders). The proportion of each of these cohorts that returned to breed at the study colony varied markedly, with no recorded recruits from some cohorts and up to 8% in others. We then used stepwise multiple regression to determine which factors influenced this observed variation in cohort recruitment rate. There was no support for any effects of the NAO, colony size or fishing activity, but there was a significant relationship between cohort recruitment rate and summer growing season temperatures in the Northern Hemisphere (Fig. 3; $F_{1,21} = 27.16$, $P < 0.001$, $r^2 = 0.56$), a measure that integrates large-scale variation in land and sea temperatures²¹.

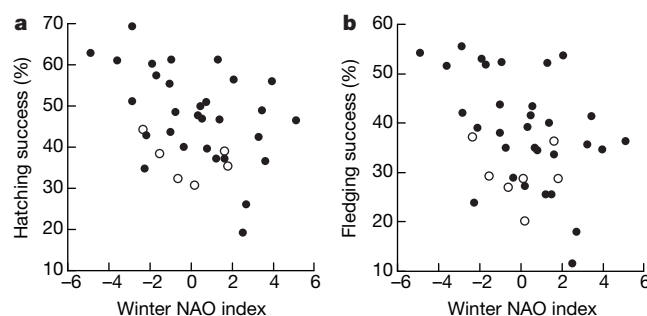


Figure 2 Variation in annual values for hatching success and fledging success in relation to the NAO index for the winter before the breeding season. **a**, Hatching success; **b**, fledging success. Data from years when there were high levels of disturbance through the incubation and nestling period (>10 days) are represented as open circles.

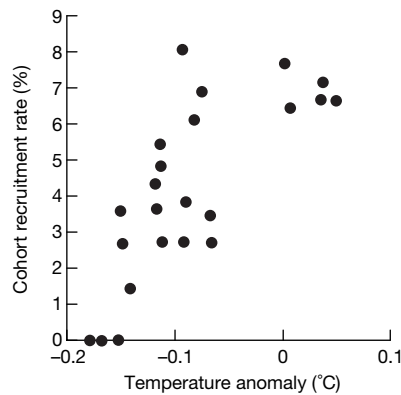


Figure 3 The recruitment rate for different cohorts of chicks from the study colony was significantly related to anomalies in Northern Hemisphere growing season temperatures during the cohort year of birth (1958–80). Recruitment rates were based on the

percentage of fledged chicks from the Eynhallow colony that were found breeding back at the colony within the following 18 years. Ninety per cent of all known-age recruits at the colony were first recorded breeding before they were 18 years old.

For many long-lived marine top predators of commercial or conservation interest, population trends are assessed through annual counts of occupied nest sites⁶, or adults present at breeding colonies^{22,23}. These findings suggest that any effects of climate change on these indices of population size will be difficult to detect without a detailed understanding of the underlying demographic processes. Given the observed effects of the winter NAO on attendance at the colony (Fig. 1) and breeding performance (Fig. 2), and of temperature on subsequent recruitment (Fig. 3), one would predict that colony size will be influenced by a complex interaction between these different climate indices. Furthermore, many long-lived vertebrates such as fulmars do not recruit until they are several years old, and the effects of changes in breeding performance and recruitment on colony dynamics should therefore be lagged.

Annual estimates of the size of our study colony show a general increase throughout the study period, but with marked oscillations over the past 30 years (Fig. 4a). As predicted, the colony size was significantly related to the previous winter's NAO index ($P < 0.03$),

and to lagged values for the winter NAO ($P < 0.05$) and Northern Hemisphere summer temperatures ($P < 0.001$), with a time lag of 5 years providing the overall best fit model ($F_{5,33} = 18.02$, $P < 0.001$, $r^2 = 0.73$). However, colony sizes were greatest after periods when recruits had been born in cooler conditions (Fig. 4b), contrary to our predictions that were based on observed effects of temperature on local recruitment (Fig. 3). These data suggest that the cooler temperatures experienced by young birds could affect cohort dispersal rates rather than survival. If so, increases in colony size after these conditions may reflect an increase in recruits from the much larger colonies found in arctic areas.

Among long-lived vertebrates, the influence of environmental variation has generally been most clearly shown through effects on reproductive success^{11,12} and early survival¹³. These findings illustrate how delayed maturity in these long-lived species will mean that the effects of climate change on breeding population size are likely to be lagged, with the period of lag depending on the species' life-history characteristics. Future attempts to detect the impacts of climate change must carefully consider contemporary population trends in relation to the effects of environmental variation on both survival and dispersal of these early life stages. □

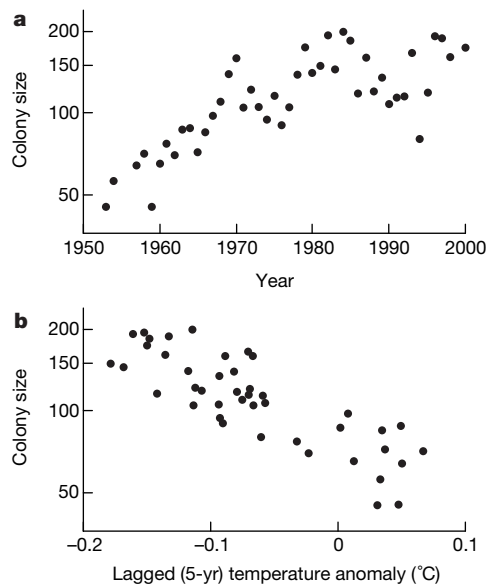


Figure 4 Variations in the number of fulmars breeding on Eynhallow. **a**, Trends in the size of the study colony (1953–2000). **b**, Changes in colony size in relation to lagged (5-year) Northern Hemisphere temperature anomalies.

Methods

Study colony and data collection

The study colony was located on a 1-km² uninhabited island in Orkney, Scotland (59° 8' N, 3° 8' W), where fulmars nest on rocky cliffs, derelict buildings, stone walls and low grassy cliffs. In most years between 1950 and 1995, visits were made in May (to search for occupied nest sites and record marked adults), July (to estimate hatching success) and August (to ring chicks)²⁴. All chicks that survived to be ringed in August were considered to have fledged. Estimates of colony size were based on the number of nests containing an egg in May. Many known nest sites remained unoccupied, even in years of peak colony size, and competition for nest sites appeared unlikely to influence recruitment rates. Each year, a sample of new recruits to the breeding population were marked with individual colour-ring combinations, and their sex was determined from bill measurements²⁵. Roughly 60% of the breeding population were marked throughout most of the study period¹⁴. Between 1996 and 2000, fewer visits were made, but these were sufficient to estimate colony size in all years except 1999, and to identify any new recruits that had been ringed as chicks.

Climate indices

Data on Northern Hemisphere temperature anomalies were based on smoothed tree ring re-constructions²¹, providing a suitably large-scale proxy that integrates marine and terrestrial growing season temperatures across the fulmar's geographical range. These data are available from the website of the National Geophysical Data Centre, USA (<http://www.ngdc.noaa.gov/paleo.html>). Data on the winter North Atlantic Oscillation index²⁶ were obtained from the website of the National Centre for Atmospheric Research, Boulder, USA (<http://www.cgd.ucar.edu/cas/catalog/climind/>). The winter NAO index is an integrated measure that influences many climatic variables such as precipitation, windspeed and temperature over large parts of the Northern Hemisphere^{19,26}. The annual Scottish Fisheries Statistics (available from Her Majesty's Stationary Office, Edinburgh)

were used to provide indices of fishing activity. These indices were based on the annual weight of landings of demersal species by Scottish boats to ports in Orkney and Shetland, and to all Scottish ports. Analyses focused on demersal catches because these fisheries were the most likely to produce by-catch and fisheries waste.

Data analyses

Estimates of cohort variation in recruitment rate were based on the 1,281 chicks that were ringed at sites on Eynhallow between 1958 and 1980. The effort put into catching the new recruits to the breeding population varied between years, but should not bias overall estimates of cohort recruitment rate because searching effort for each cohort was integrated over 18 field seasons. The probability of observing breeding adults at the colony was estimated for colour-ringed birds that were confirmed breeders and known to be alive in that year (on the basis of positive sightings in subsequent years). Analyses were based on observations made between 1954 and 1995, using a total sample of 987 colour-ringed breeding adults.

To standardize effort across years, estimates of colony size, hatching success and fledging success excluded data from part of the colony where steep cliffs prevented easy access to nest sites. The sub-area used in this study represented roughly 64% of the whole colony, and these estimates therefore differ slightly from previously published data from all nests²⁴. A highly significant relationship between the number of occupied nests in this sub-area and the whole colony (data from 1950 to 1995: $F_{1,44} = 808$, $P < 0.0001$, $r^2 = 0.95$) suggests that its demographic development was representative.

Because disturbance has been shown previously to influence breeding performance²⁷, we also incorporated this factor into our analysis of variations in reproductive success. Annual estimates of disturbance levels were based on the number of days that researchers were present on the island during the incubation and chick-rearing period. The influence of different climate indices, fishing effort and disturbance on cohort recruitment rate and breeding performance (hatching success and fledging success) was determined by stepwise multiple regression. Initial factors included in each model were (1) colony size; (2) Northern Hemisphere temperature anomaly; (3) winter NAO; (4) annual NAO; (5) year; (6) Scottish demersal catch; (7) Orkney and Shetland demersal catch; and (8) disturbance.

The demographic consequences of observed environmentally induced variation in reproductive success and recruitment were initially analysed using an autoregressive model that estimated the influence of the following variables on subsequent colony size: (1) winter NAO index (which is predicted to influence attendance; Fig. 1); (2) lagged values for colony size; and (3) lagged values for the winter NAO index (which are predicted to influence the number of fledglings produced in previous cohorts; Fig. 2b); and (4) lagged values for the Northern Hemisphere temperatures (which are predicted to influence the subsequent recruitment rate for each cohort of fledglings; Fig. 3). The effect of lagged values for colony size was not significant, and this variable was therefore dropped from the final model. Year was included in each model to account for the strong temporal trend in the data. Values of 1–9 years were used for the time lag in the effects of colony size and climate indices on future colony size, on the basis of biologically reasonable estimates of the delay between fledging and first reproduction¹⁴.

Received 25 June; accepted 24 July 2001.

- Post, E., Peterson, R. O., Stenseth, N. C. & McLaren, B. E. Ecosystem consequences of wolf behavioural response to climate. *Nature* **401**, 905–907 (1999).
- Milner, J. M., Elston, D. A. & Albon, S. D. Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *J. Anim. Ecol.* **68**, 1235–1247 (1999).
- Sæter, B.-E. *et al.* Population dynamical consequences of climate change for a small temperate songbird. *Science* **287**, 854–856 (2000).
- Planque, B. & Taylor, A. H. Long-term changes in zooplankton and the climate of the North Atlantic. *ICES J. Mar. Sci.* **55**, 644–654 (1998).
- O'Brien, C. M., Fox, C. J., Planque, B. & Casey, J. Climate variability and North Sea cod. *Nature* **404**, 142 (2000).
- Lloyd, C., Tasker, M. L. & Partridge, K. *The Status of Seabirds in Britain and Ireland* (T & A. D. Poyser, London, 1991).
- Fisher, J. *The Fulmar* (Collins, London, 1952).
- Salomonsen, E. The geographical variation of the fulmar (*Fulmar glacialis*) and the zones of marine environment in the North Atlantic. *Auk* **82**, 327–355 (1965).
- Brown, R. G. B. Fulmar distribution: a Canadian perspective. *Ibis* **112**, 44–51 (1970).
- Wynne-Edwards, V. *Animal Dispersion in Relation to Social Behaviour* (Oliver & Boyd, Edinburgh, 1962).
- Kitaysky, A. S. & Golubova, E. G. Climate change causes contrasting trends in reproductive performance of planktivorous and piscivorous alcids. *J. Anim. Ecol.* **69**, 248–262 (2000).
- Post, E. & Stenseth, N. Climatic variability, plant phenology, and northern ungulates. *Ecology* **80**, 1322–1339 (1999).
- Gaillard, J.-M., Festa-Bianchet, M. & Yoccoz, N. G. Population dynamics of large herbivores: variable recruitment with constant adult survival. *Trends Ecol. Evol.* **13**, 58–63 (1998).
- Ollason, J. C. & Dunnet, G. M. in *Reproductive Studies of Individual Breeding Systems* (ed. Clutton-Brock, T. C.) 263–278 (Univ. Chicago Press, Chicago, 1988).
- Dunnet, G. M. Population studies of the Fulmar on Eynhallow, Orkney Islands. *Ibis* **133**, 24–27 (1991).
- Mougin, J. L., Jouanin, C. & Roux, F. Intermittent breeding in Cory's Shearwater *Calonectris diomedea* of Selvagem Grande, North Atlantic. *Ibis* **139**, 40–44 (1997).
- Bradley, J. S., Wooller, R. D. & Skira, I. J. Intermittent breeding in the short-tailed shearwater *Puffinus tenuirostris*. *J. Anim. Ecol.* **69**, 639–650 (2000).
- Furness, R. W. & Bryant, D. M. Effect of wind on field metabolic rates of breeding Northern Fulmars. *Ecology* **77**, 1181–1188 (1996).

- Wilby, R. L., O'Hare, G. & Barnsley, N. The North Atlantic Oscillation and British Isles climate variability, 1965–1996. *Weather* **52**, 266–276 (1997).
- Phillips, R. A. *et al.* Diet of the northern fulmar *Fulmar glacialis*; reliance on commercial fisheries? *Mar. Biol.* **135**, 159–170 (1999).
- Briffa, K. R., Jones, P. D., Schweingruber, F. H. & Osborn, T. J. Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature* **393**, 450–455 (1998).
- Cooper, C. F. & Stewart, B. S. Demography of Northern elephant seals, 1911–1982. *Science* **219**, 969–971 (1983).
- Pomeroy, P. P., Twiss, S. D. & Duck, C. D. Expansion of a grey seal (*Halichoerus grypus*) breeding colony: changes in pupping site use at the Isle of May, Scotland. *J. Zool.* **250**, 1–12 (2000).
- Dunnet, G. M., Ollason, J. C. & Anderson, A. A 28-year study of breeding Fulmars *Fulmar glacialis* in Orkney. *Ibis* **121**, 293–300 (1979).
- Dunnet, G. M. & Anderson, A. A method of sexing living Fulmars in the hand. *Bird Study* **8**, 119–126 (1961).
- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Ollason, J. C. & Dunnet, G. M. Nest failures in the Fulmar: the effect of observers. *J. Field Ornithol.* **51**, 39–54 (1980).

Acknowledgements

We are indebted to the late G. Dunnet, who developed and directed the long-term field study on which these findings are based. Thanks also to A. Anderson, P. Heppleston, P. Doyle, P. Cosgrove and many other colleagues who assisted with data collection and analyses; and to S. D. Albon, P. F. Billingsley, J. W. Durban, D. A. Elston, M. L. Gorman, X. Lambin, J. G. Ollason and W. R. Turrell for advice and discussion. Orkney Islands Council and the late J. M. Robertson kindly provided access to facilities on Eynhallow.

Correspondence and requests for materials should be addressed to P.M.T. (e-mail: lighthouse@abdn.ac.uk).

Histamine regulates T-cell and antibody responses by differential expression of H1 and H2 receptors

Marek Jutel*, Takeshi Watanabe†, Sven Klunker*, Mübcecel Akdis*, Olivier A. R. Thomet*, Jozef Malolepszy‡, Teresa Zak-Nejmark‡, Ritsuko Koga†, Takashi Kobayashi†, Kurt Blaser* & Cezmi A. Akdis*

* Swiss Institute of Allergy and Asthma Research (SIAF), Obere Strasse 22, CH-7270 Davos, Switzerland

† Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, 812-8582 Fukuoka, Japan

‡ Department of Internal Medicine and Allergology, Wrocław Medical University, Traugutta 57, 50-417 Wrocław, Poland

Many pathological processes, including those causing allergies and autoimmune diseases, are associated with the presence of specialized subsets of T helper cells (T_H1 and T_H2) at the site of inflammation^{1–4}. The diversity of T_H1 and T_H2 function is not predetermined but depends on signals that drive the cells towards either subset^{1–4}. Histamine, released from effector cells (mast cells and basophils) during inflammatory reactions can influence immune response^{5–8}. Here we report that histamine enhances T_H1 -type responses by triggering the histamine receptor type 1 (H1R), whereas both T_H1 - and T_H2 -type responses are negatively regulated by H2R through the activation of different biochemical intracellular signals. In mice, deletion of H1R results in suppression of interferon (IFN)- γ and dominant secretion of T_H2 cytokines (interleukin (IL)-4 and IL-13). Mutant mice lacking H2R showed upregulation of both T_H1 and T_H2 cytokines. Relevant to T-cell cytokine profiles, mice lacking H1R displayed increased specific antibody response with increased immunoglobulin- ϵ (IgE) and IgG1, IgG2b and IgG3 compared with mice lacking H2R. These findings account for an important regulatory mechanism in the control of inflammatory functions through effector-cell-derived histamine.

H1R and H2R both belong to the G-protein-coupled receptor family^{9,10}; however, they trigger different intracellular events upon activation^{5,6,11}. We investigated the expression of these receptors on T_H1 and T_H2 cells bearing the CD4 antigen (CD4⁺) and whether histamine stimulation exerts differential effects on these cells. Cells bearing the histamine receptor were stained with fluorescein-labelled histamine; histamine bound significantly more strongly to T_H1 cells than to T_H2 cells (Fig. 1a). The H1R-specific antagonist tripeleennamine inhibited histamine binding in T_H1 but not in T_H2 cells, which indicates predominant H1R expression on T_H1 cells. Neither the H2R antagonist ranitidine nor the H3R antagonist clobenpropit had any effect on histamine binding to T_H1 cells (not shown). IL-3 significantly increased histamine receptor expression on T_H1 but not on T_H2 cells (Fig. 1b, c). Predominant expression of H1R on T_H1 cells and H2R on T_H2 cells is demonstrated by antibodies generated against the H1R and H2R (Fig. 1d, e). Predominant expression of H1R messenger RNA was detected in T_H1 cells compared with T_H2 cells (Fig. 1f). Furthermore, IL-3 enhanced expression of H1R mRNA in T_H1 but not in T_H2 cells. In contrast, H2R mRNA was predominantly expressed in T_H2 cells (Fig. 1g).

We then investigated the changes in histamine receptor expression during naive cell differentiation towards T_H1 and T_H2 cells. CD45RA⁺ T cells expressed little or undetectable mRNA and no detectable HRs on the surface (Fig. 1a, h, i). Stimulation of CD45RA⁺ T cells with IL-12 resulted in enhanced H1R expression, whereas IL-4 suppressed H1R mRNA expression (Fig. 1h, i). These results demonstrate that CD45RO⁺ T_H1 and T_H2 cells preferentially but not exclusively express H1R and H2R, respectively, and that H1R and H2R are regulated by cytokines present in the immune environment.

The differential expression of H1R and H2R on T_H1 and T_H2 cells should result in distinct intracellular events triggered upon receptor stimulation. Therefore, we investigated the signals induced by histamine in T_H1 and T_H2 cells. The primary signal from H1R is delivered through calcium-dependent activation of phospholipase C with generation of inositol 1,4,5-triphosphate (IP₃) and 1,2-diacylglycerol (DAG)¹¹. Because H1R is overexpressed on T_H1 cells, stimulation of T_H1 (Fig. 2a) but not T_H2 cells (Fig. 2b) with histamine resulted in significant Ca²⁺ flux. This effect was specifically blocked by tripeleennamine, demonstrating H1R-dominated signalling in T_H1 cells. Activation of H1R can produce signals

via the Gq/11 family of G proteins, which are capable of communicating with signalling mediated by T-cell receptors. The Gq-mediated signal induces translocation of the nuclear factor (NF)-AT to the nucleus, which is involved in activation of the IL-2 gene¹². Stimulation through H2R results in cAMP formation^{5,6,13}. T_H2 cells are characterized by high levels of cyclic AMP¹⁴. The cAMP-inducible transcriptional repressor protein ICER (inducible cAMP early repressor), which inhibits activity of the IL-2 promoter, has been described in T cells¹⁵. Accordingly, only T_H2 cells showed dose-dependent increases in cAMP to histamine stimulation (Fig. 2c). This effect was fully abrogated by the specific H2R antagonist ranitidine but not by tripeleennamine (Fig. 2d), indicating the H2R-dominated signalling in T_H2 cells.

We further investigated whether distinct intracellular events induced in T_H1 and T_H2 cells on stimulation with histamine influence the functions of immune effectors. We analysed the effect of histamine on proliferation and cytokine secretion of human T_H1 and T_H2 cells. Histamine enhanced anti-CD3-induced proliferation of T_H1 cells, whereas T_H2 cells were inhibited (Fig. 3a). Tripeleennamine fully blocked the effect of histamine on T_H1 cells, whereas ranitidine blocked the action of histamine on T_H2 cells. The relevance of these findings was investigated on T_H1- and T_H2-type antigen-specific responses. Peripheral blood mononuclear cells (PBMCs) from sensitized individuals were stimulated with either purified protein derivative of *Mycobacterium bovis* (PPD) or allergens (Der p1, a major house dust mite allergen, and phospholipase A₂ (PLA), a major bee venom allergen). Predominance of T_H1 has been reported in PPD-specific T-cell responses¹⁶, and histamine enhanced proliferation in PPD-stimulated cultures. In contrast, allergen-stimulated cells, which showed T_H2 predominance, from individuals allergic to house dust mites¹⁷ and bee venom¹⁸ were suppressed (Fig. 3b). We also observed that secretion of T_H1 cytokine IFN-γ was enhanced in T_H1 cells, whereas T_H2 cytokines (IL-4, IL-13) were inhibited in T_H2 cells by histamine (Fig. 3c).

To provide direct evidence linking our findings to *in vivo* immunoregulatory processes, we investigated mutant mice lacking H1R and H2R. Both types of mice developed normally and were apparently healthy and fertile^{19,20}. It has been demonstrated that deletion of either H1R or H2R results in abnormalities in the central nervous and gastrointestinal systems^{19–21}. Here we show that H1R- and H2R-deleted mice are characterized by a marked deviation of the immune response. H1R- and H2R-deleted mice showed

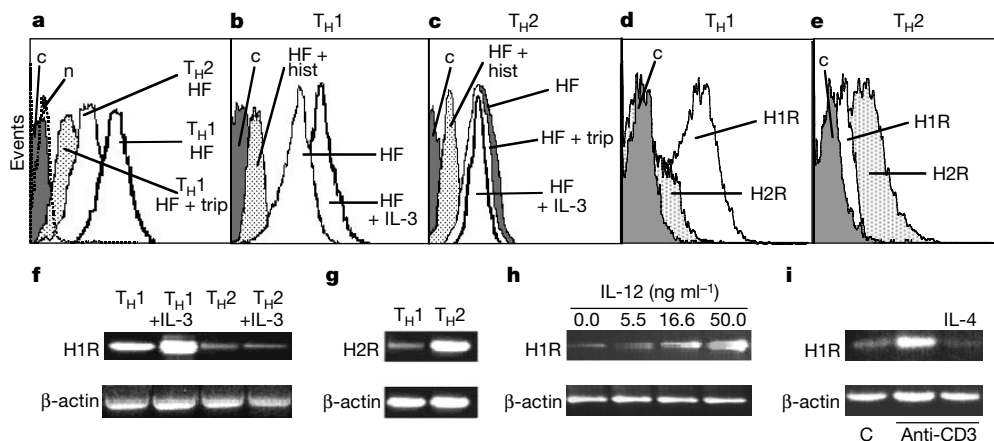


Figure 1 Histamine receptor expression on T_H1 and T_H2 cells. **a**, Inhibition of histamine fluorescein (HF) binding by H1R antagonist tripeleennamine (trip) in T_H1 but not T_H2 cells (c, control FITC; n, CD45RA⁺ T cells). **b**, Specificity of HF cell binding confirmed by inhibition with unlabelled histamine (hist). IL-3 increases H1R expression on T_H1 cells. **c**, HR expression on T_H2 cells is neither enhanced by IL-3 nor inhibited by tripeleennamine. **d**, **e**, Preferential but not exclusive H1R and H2R expression on T_H1 and T_H2 cells,

respectively, by specific antibodies. **f**, T_H1 cells show high H1R mRNA expression. IL-3 stimulation results in enhanced H1R mRNA on T_H1 but not on T_H2 cells. **g**, T_H2 cells express high H2R mRNA. **h**, IL-12 induces increased H1R mRNA in CD45RA⁺ T cells. **i**, IL-4 suppresses H1R mRNA expression in CD45RA⁺ T cells. Results are of six flow cytometric measurements and three mRNA determinations from different donors.

differences in development of T_H1 and T_H2 cells, as demonstrated by frequencies of IL-4- and IFN- γ -producing spleen T cells (Fig. 4a). H1R-deleted mice showed lower, whereas H2R-deleted mice showed higher, percentages of IFN- γ -producing T cells compared with wild-type mice ($P < 0.001$). A tendency for increased frequency of IL-4-producing cells was observed both in H1R- and H2R-deleted mice compared with wild-type mice. After histamine stimulation, consistent with the findings in humans, H1R-deleted mice demonstrated abundant suppression of IFN- γ secretion by T cells (Fig. 4b). In contrast, the T_H2 cytokines (IL-4 and IL-13) were enhanced.

These results demonstrate the importance of H1R-mediated signalling in the induction of T_H1 -type cytokine responses. It was previously reported that the signal generated by histamine through H1R augments immune responses mediated by the antigen receptor, suggesting a cross-talk between H1R and antigen-receptor-mediated signalling⁸. T-cell proliferation and ZAP-70 tyrosine kinase phosphorylation were strongly reduced in H1R-deleted mice⁸. H2R-deleted mice showed significant enhancement of both T_H1 - and T_H2 -type cytokine secretion by histamine pre-incubation (Fig. 4b). This indicates that H2R triggering provides a negative signal for T cells. Activation of H2R on T cells increases intracellular cAMP and thus downregulates both T_H1 - and T_H2 -type cytokines.

The influence of histamine and differential HR expression on T_H1 and T_H2 cells was further analysed in ovalbumin (OVA)-specific, T-cell-dependent antibody production in H1R- and H2R-deleted mice (Fig. 5). H1R-deleted mice produced high OVA-specific IgG1 and IgE compared with wild-type mice ($P < 0.05$). In contrast, H2R-deleted mice showed decreased serum levels of OVA-specific IgG3 and IgE compared with wild-type mice ($P < 0.01$). H1R-deleted mice produced significantly higher amounts of OVA-specific IgE, IgG1, IgG2b and IgG3 compared with H2R-deleted mice ($P < 0.001$). These results demonstrate the dominant role of H1R and related T_H1 response in the suppression of humoral

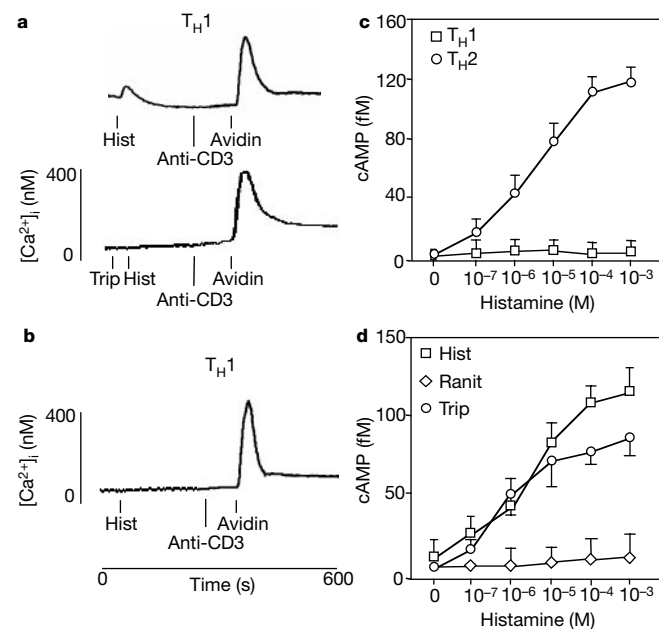


Figure 2 Histamine-induced intracellular calcium $[Ca^{2+}]_i$ flux and cAMP in T_H1 and T_H2 cells. **a**, $[Ca^{2+}]_i$ flux in histamine (hist) stimulated T_H1 cells and blockage by tripelenamine (trip). Crosslinking of the CD3 molecule was used as control. **b**, Histamine does not induce $[Ca^{2+}]_i$ flux in T_H2 cells. **c**, Histamine-induced intracellular cAMP levels in T_H2 but not in T_H1 cells. **d**, Histamine-induced cAMP in T_H2 cells was inhibited by ranitidine (ranit) but not tripelenamine. Mean $\pm$ s.d. of triplicates are shown. The data are representative of four separate experiments performed under identical conditions.

immune response. Accordingly, IgG2a, an isotype dependent on the T_H1 cytokine IFN- γ , did not show any suppression in H2R-deleted mice. H2R-deleted mice showed significantly decreased OVA-specific IgE production compared with wild-type and H1R-deleted mice. Although T cells of H2R-deleted mice secreted increased IL-4 and IL-13, OVA-specific IgE was suppressed in the presence of highly increased IFN- γ , supporting the dominant role of IFN- γ on the suppression of IgE production²².

Collectively, our results indicate that histamine can modulate T-cell-mediated immune response. Mast cells and basophils, which are considered mainly as effector cells for IgE-mediated reactions, can modulate the immune response through secreted histamine^{23,24}. IL-3 is important in the regulation of mast cell and basophil development by T cells²⁵. The increase in H1R expression on T_H1 cells by IL-3 may account for negative-feedback regulation due to an anti-allergic role of T_H1 cells. We conclude that histamine secreted from effector cells of inflammation potently

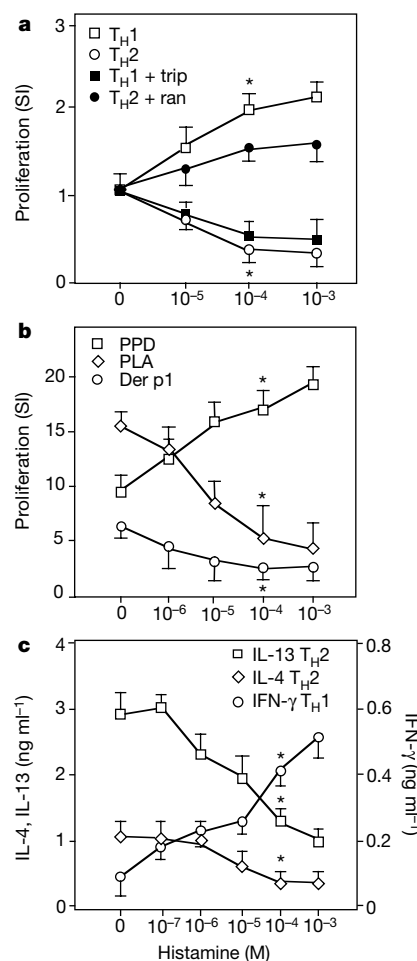


Figure 3 Regulation of T_H1 - and T_H2 -type responses by histamine. **a**, Influence of histamine on anti-CD3-induced proliferation of T_H1 and T_H2 cells; blockage by tripelenamine (trip) and ranitidine (ran). Stimulation index (SI) of 1 representing proliferation with anti-CD3 in the absence of histamine corresponds to 51,000 c.p.m. in T_H1 cells and 35,000 c.p.m. in T_H2 cells. **b**, The influence of histamine on T_H1 - and T_H2 -type proliferative responses to antigens in PBMC cultures. SI of 1 representing background proliferation of unstimulated cells was 1,600 c.p.m. in PPD, 1,100 c.p.m. in PLA and 1,300 c.p.m. in Der-p1-stimulated cultures. **c**, Effect of histamine on cytokine secretion in anti-CD3-stimulated T_H1 and T_H2 cells. Results shown are representative of six experiments performed under identical conditions. Mean $\pm$ s.d. of triplicates are shown (asterisk, $P < 0.001$, paired t -test between 10^{-4} M histamine dose and without histamine; the median effective dose ED_{50} was 5×10^6 M).

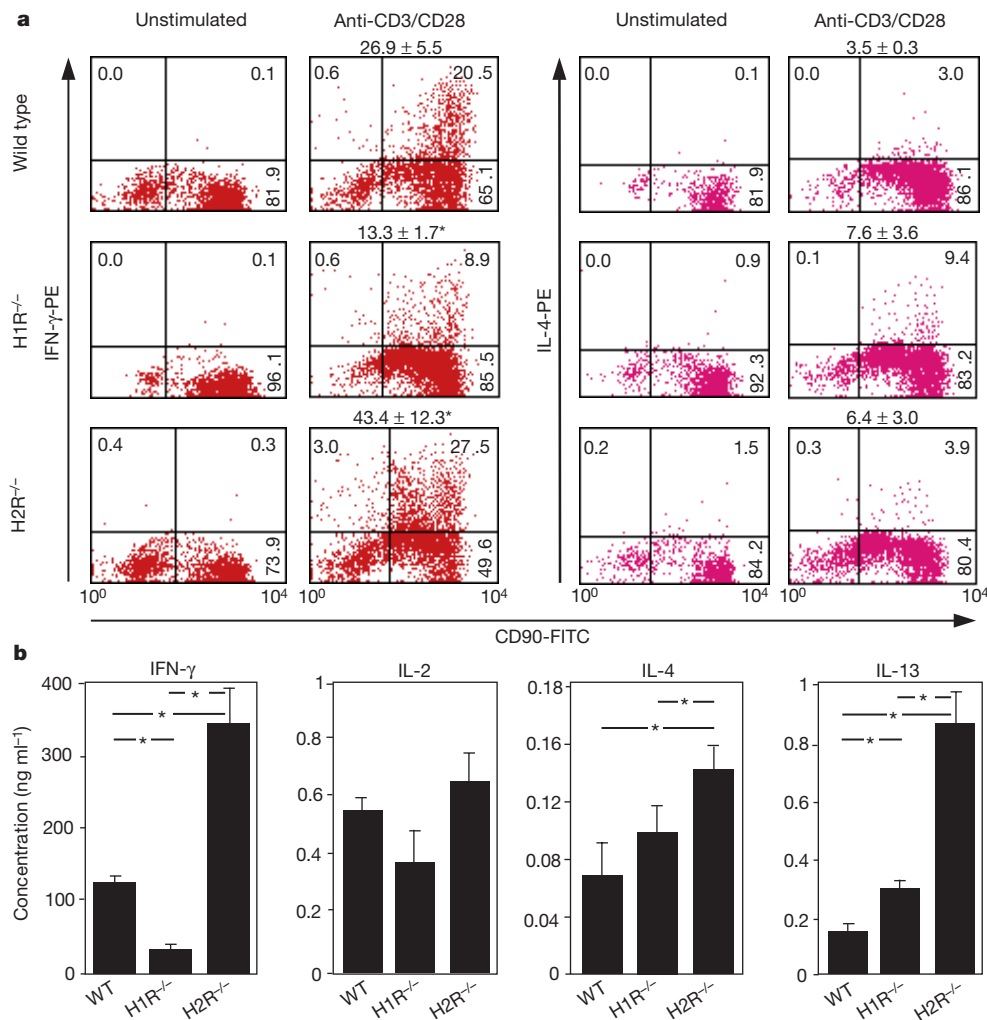
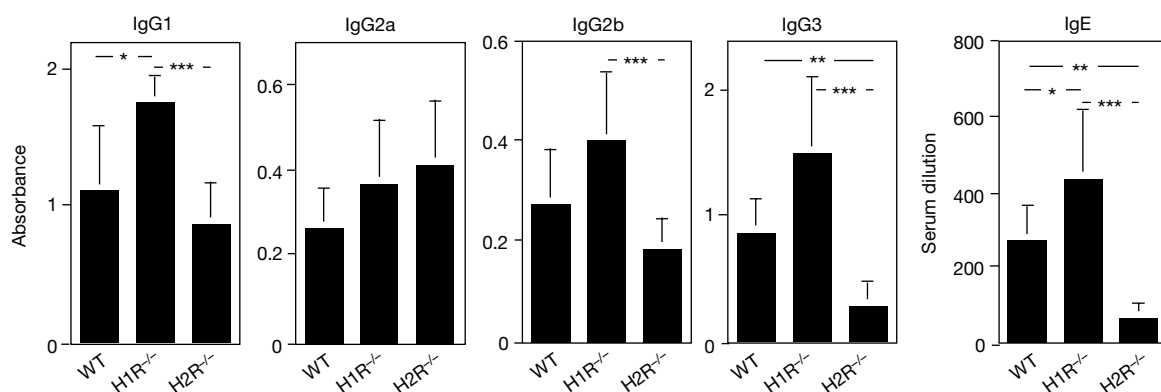


Figure 4 T-cell cytokine profiles in wild-type (WT), H1R-deleted and H2R-deleted mice. **a**, Spleen T cells were stimulated with anti-CD3 and anti-CD28 antibodies and intracytoplasmic IL-4 and IFN-γ were detected after 48 h. Cells were counter-stained with anti-CD90 antibody. The percentage of positive cells is shown on each quadrant. Data are representative of six mice in each group. Mean ± s.d. of cytokine-positive T cells is depicted over each histogram. The same changes in cytokine profiles were observed after

stimulation for 8 h and 24 h (data not shown). **b**, Histamine-pretreated spleen cells were stimulated with anti-CD3 antibodies and cytokines were detected from supernatants collected after 72 h. The results shown are mean ± s.d. of triplicate cultures from three mice of each type. The same results were obtained in six experiments (asterisk, $P < 0.001$).



100,000-fold for IgG1, 500-fold for IgG2a, 50,000-fold for IgG2b and 500-fold for IgG3. IgE was determined by passive cutaneous anaphylaxis. Results demonstrate mean ± s.d. of two experiments each with six mice. Triple asterisk, $P < 0.001$; double asterisk, $P < 0.01$; single asterisk, $P < 0.05$.

influences T_H1 and T_H2 responses as well as antibody isotypes as a regulatory loop in inflammatory reactions. □

Methods

H1R- and H2R-deleted mice

H1R- and H2R-deleted mice were generated and the mutant mice were backcrossed to C57BL/6 mice six and eight times, respectively, and maintained under pathogen-free conditions with proper aeration, light and temperature^{19,20}.

Differentiation of human T_H1 and T_H2 cells

T_H1 and T_H2 cells were obtained by differentiation of CD45RA⁺ T cells derived from cord blood, as described before³. Cytokine production by T_H1 and T_H2 cells revealed distinct profiles. T_H1 cells (10^6 per ml) secreted 5.6 ng ml⁻¹ of IFN- γ , <0.1 ng ml⁻¹ of IL-4, <0.1 ng ml⁻¹ of IL-5 and <0.2 ng ml⁻¹ of IL-13. T_H2 cells (10^6 per ml) secreted 4.6 ng ml⁻¹ of IL-4, 7.2 ng ml⁻¹ of IL-5, 9.2 ng ml⁻¹ of IL-13, and <0.1 ng ml⁻¹ of IFN- γ when given $10 \mu\text{g ml}^{-1}$ plate-bound anti-CD3 stimulation in 48 h.

Flow cytometry and mRNA expression

For flow cytometric analysis, 5×10^4 cells were stained either with fluorescein-conjugated histamine (Molecular Probes) alone or together with histamine dihydrochloride, H1R antagonist triplenamine hydrochloride, H2R antagonist ranitidine dihydrochloride or H3R antagonist clobenpropit dihydrobromide (RBI Research Biochemicals). Mouse IgG was used as a negative control. Fluorescein-conjugated histamine (10^{-3} M), histamine (10^{-3} M), triplenamine (10^{-6} M), ranitidine (6.3×10^{-3} M) and clobenpropit (1.25×10^{-7} M) were used at optimized concentrations. The concentrations of triplenamine, ranitidine and clobenpropit were 1,000 times the pA_2 value, that is, the negative logarithm of the concentration required to get 50% reduction of the efficacy of histamine¹¹. The specificity of histamine receptor staining was further demonstrated by rabbit polyclonal antibodies generated against the amino-terminal peptides of H1R (SCLEDKMCEN) and H2R (SFCLDSTACKIT).

For the semiquantitative analysis by polymerase chain reaction with reverse transcription (RT-PCR) of H1R and H2R, 2×10^6 T_H1 , T_H2 or purified CD45RA⁺ T cells were cultured in RPMI-1640 medium^{26,27} with or without 25 ng ml⁻¹ IL-3 (PharMingen), 25 ng ml⁻¹ IL-4 (Novartis), or 5.5, 16.6 or 50.0 ng ml⁻¹ IL-12 (PharMingen). Solid-phase bound anti-CD3 monoclonal antibody (Ortho Diagnostics) was used at $10 \mu\text{g ml}^{-1}$. Total RNA was isolated using Trizol reagent (Life Technologies) following the instructions of the manufacturer. One step RT-PCR was performed using the One Tube Titan RT-PCR System (Roche). Primers used to amplify specific gene products were as follows: H1R forward primer, 5'-ttt ctc tct ttt ctg tgg gtt att c-3'; H1R reverse primer, 5'-ctt ggg ggt ttg gga tgg tga ctt c-3'; H2R forward primer, 5'-ttc atc gtg tcc ttg gct atc-3'; H2R reverse primer, 5'-ctg aga ggc gtt gga cct c-3'; β -actin forward primer, 5'-atc tgc cac cac acc ttc tac aat gag ctg cg-3'; β -actin reverse primer, 5'-cgt cat act cct gct tgc tga tcc aca tct gc-3'. H1R PCR and H2R PCR products were 387 base pairs (bp) and 804 bp, respectively. β -actin was used as the housekeeping gene. Primers amplified a 838-bp PCR product. The reverse transcription of RNA was performed, 50°C for 30 min, followed by 94°C for 2 min. Conditions for PCR were 10 cycles of 94°C for 30 s, 55°C for H1R (62°C for H2R) for 30 s (annealing), 68°C for 45 s (elongation), followed by 25 cycles of 94°C for 30 s and 55°C for H1R (62°C for H2R) for 30 s (annealing) and 68°C for 45 s plus 5 s per step (elongation). The elongation time was prolonged in one cycle to 7 min (68°C). PCR products were identified by 1% agarose gel electrophoresis.

Intracellular calcium and cAMP measurements

The intracellular ionized calcium concentration ($[Ca^{2+}]_i$) was measured by a spectrofluorometric assay as previously described²⁸. T_H1 or T_H2 cells (4×10^6), labelled with fura-2/AM, were stimulated with histamine 5×10^{-3} M or with biotinylated monoclonal antibody to CD3 (Sigma) ($7 \mu\text{g ml}^{-1}$) followed by crosslinking with Avidin (Sigma) ($1 \mu\text{g ml}^{-1}$). Triplenamine and ranitidine were used at 10^{-6} M. Intracellular cAMP levels were determined using the Biotrak cAMP enzyme immunoassay system (Amersham Pharmacia) according to the manufacturer's instructions.

Cell cultures and cytokine detection

Human T_H1 and T_H2 cells were incubated for 60 min with histamine either alone or with H1R antagonist triplenamine (10^{-6} M) or with H2R antagonist ranitidine (10^{-6} M). PBMCs were incubated with increasing concentrations of histamine for 2 h. Thereafter 10% FCS was supplemented and the cells were stimulated with $10 \mu\text{g ml}^{-1}$ solid-phase bound anti-CD3 monoclonal antibodies in triplicate. Cytokine levels were measured by ELISA after 48 h^{26,27}. Incorporation of [³H]-thymidine was determined after 48 h^{26,27}. PBMCs were stimulated with either Der p1 ($10 \mu\text{g ml}^{-1}$), PLA ($10 \mu\text{g ml}^{-1}$) or PPD ($2 \mu\text{g ml}^{-1}$)²³. Incorporation of [³H]-thymidine was measured in antigen-specific cultures after 5 d.

Spleen T cells from 8-week-old wild-type, H1R-deleted and H2R-deleted mice were purified by passing them through a nylon wool column, and red blood cells were lysed with NH₄Cl lysis buffer. The cells were then suspended in RPMI-1640 medium. Intracytoplasmic IL-4 and IFN- γ were determined after 8, 24 and 48 h of stimulation with $10 \mu\text{g ml}^{-1}$ immobilized anti-CD3e and $10 \mu\text{g ml}^{-1}$ soluble anti-CD28 antibody (both from PharMingen). Monensin (2 μM , Sigma) was added 4 h before the end of cultures. Cells were fixed, permeabilized, stained with PE-labelled anti-IFN- γ and PE-labelled anti-IL-4 (ref. 27). Because CD3 was downregulated during anti-CD3 stimulation, anti-CD90

labelled with fluorescein isothiocyanate (FITC; Thy1.2, PharMingen) was used as a pan T-cell marker for counter stain. Spleen T cells were incubated with 10 mM histamine in the absence of FCS at 37°C for 2 h. After being supplemented with 10% FCS, T cells (2×10^6 per ml) were stimulated with $10 \mu\text{g ml}^{-1}$ solid-phase bound anti-CD3e antibody in triplicate in 24-well plates. Cytokine levels in the 3-d supernatants were determined by ELISA (Amersham Pharmacia).

Determination of OVA-specific antibody production

Eight- to nine-week-old wild-type, H1R-deleted and H2R-deleted mice were immunized intraperitoneally with 100 μg OVA adsorbed to aluminium hydroxide (both from Sigma) at days 0 and 21. Serum was collected at day 28 and OVA-specific IgG1, IgG2a, IgG2b and IgG3 were determined in serial dilutions in duplicates by ELISA⁸. OVA-specific IgE was determined by passive cutaneous anaphylaxis as described²⁹.

Received 22 May; accepted 13 August 2001.

- Abbas, A. K., Murphy, K. M. & Sher, A. Functional diversity of helper T lymphocytes. *Nature* **383**, 787–793 (1996).
- Chen, Q. *et al.* Development of Th1-type immune responses requires the type 1 cytokine receptor TCCR. *Nature* **407**, 916–920 (2000).
- Sallusto, F., Mackay, C. R. & Lanzavecchia, A. Selective expression of the eotaxin receptor CCR3 by human T helper 2 cells. *Science* **277**, 2005–2007 (1997).
- Mosmann, T. R., Cherwinski, H., Bond, M. W., Giedlin, M. A. & Coffman, R. L. Two types of murine helper T cell clones. I. Definition according to profiles of lymphokine activities and secreted proteins. *J. Immunol.* **136**, 2348–2357 (1986).
- Roszkowski, W., Plaut, M. & Lichtenstein, L. M. Selective display of histamine receptors on lymphocytes. *Science* **195**, 683–685 (1977).
- Beer, D. J. & Rocklin, R. E. Histamine modulation of lymphocyte biology: Membrane receptors, signal transduction, and functions. *Crit. Rev. Immunol.* **7**, 55–91 (1987).
- Vannier, E. & Dinarello, C. Histamine enhances interleukin (IL)-1-induced IL-1 gene expression and protein synthesis via H2 receptors in peripheral blood mononuclear cells. Comparison with IL-1 receptor antagonist. *J. Clin. Invest.* **92**, 281–287 (1993).
- Banu, Y. & Watanabe, T. Augmentation of antigen receptor-mediated responses by histamine H1 receptor signaling. *J. Exp. Med.* **189**, 673–682 (1999).
- Yamashita, M. *et al.* Expression cloning of a cDNA encoding the bovine histamine H1 receptor. *Proc. Natl Acad. Sci. USA* **88**, 11515–11519 (1991).
- Gantz, I. *et al.* Molecular basis for the interaction of histamine with the histamine H2 receptor. *J. Biol. Chem.* **267**, 20840–20843 (1992).
- Leurs, R., Smit, M. J. & Timmerman, H. Molecular pharmacological aspects of histamine receptors. *Pharmacol. Ther.* **66**, 413–463 (1995).
- Boss, V., Talpade, D. J. & Murphy, T. J. Induction of NFAT-mediated transcription by Gq-coupled receptors in lymphoid and non-lymphoid cells. *J. Biol. Chem.* **271**, 10429–10432 (1996).
- Plaut, M., Marone, G. & Gillespie, E. The role of cyclic AMP in modulating cytotoxic T lymphocytes. II. Sequential changes during culture in responsiveness of cytotoxic lymphocytes to cyclic AMP-active agents. *J. Immunol.* **131**, 2945–2952 (1983).
- Novak, T. J. & Rothenberg, E. V. cAMP inhibits induction of interleukin 2 but not of interleukin 4 in T cells. *Proc. Natl Acad. Sci. USA* **87**, 9353–9357 (1990).
- Bodor, J., Spetz, A., Strominger, J. L. & Habener, J. F. cAMP inducibility of transcriptional repressor ICER in developing and mature T lymphocytes. *Proc. Natl Acad. Sci. USA* **93**, 3536–3541 (1996).
- Del Prete, G. F., De Carli, M. & Mastromaro, C. Purified protein derivative of *Mycobacterium tuberculosis* and excretory-secretory antigen(s) of *Toxocara canis* expand *in vitro* human T cells with stable and opposite (type 1 T helper or type 2 T helper) profiles of cytokine production. *J. Clin. Invest.* **88**, 344–350 (1991).
- Robinson, D. S. *et al.* Predominant Th2-like bronchoalveolar T lymphocyte population in atopic asthma. *N. Engl. J. Med.* **326**, 298–304 (1992).
- Akdis, C. A., Joss, A., Akdis, M., Faith, A. & Blaser, K. A molecular basis for T cell suppression by IL-10: CD28-associated IL-10 receptor inhibits CD28 tyrosine phosphorylation and phosphatidylinositol 3-kinase binding. *FASEB J.* **14**, 1666–1668 (2000).
- Inoue, I. *et al.* Impaired locomotor activity and exploratory behaviour in mice lacking histamine H1 receptors. *Proc. Natl Acad. Sci. USA* **93**, 13316–13320 (1996).
- Kobayashi, T. *et al.* Abnormal function and morphological regulation of the gastric mucosa in histamine H2 receptor-deficient mice. *J. Clin. Invest.* **105**, 1741–1749 (2000).
- Mobarakeh, I. J. *et al.* Role of histamine H(1) receptor in pain perception: a study of the receptor gene knockout mice. *Eur. J. Pharmacol.* **391**, 81–89 (2000).
- Corry, D. B. & Kheradmand, F. Induction and regulation of the IgE response. *Nature* **402**, B18–B23 (1999).
- Akdis, C. A. & Blaser, K. IL-10 induced peripheral T cell anergy and reactivation by micro-environmental cytokines: two key steps in specific immunotherapy. *FASEB J.* **13**, 603–609 (1999).
- Jutel, M. *et al.* Influence of bee venom immunotherapy on degranulation and leukotriene generation in human blood basophils. *Clin. Exp. Allergy* **26**, 1112–1118 (1996).
- Lantz, C. S. *et al.* Role for interleukin-3 in mast-cell and basophil development and in immunity to parasites. *Nature* **392**, 90–93 (1998).
- Akdis, C. A., Blesken, T., Akdis, M., Wuthrich, B. & Blaser, K. The role of IL-10 in specific immunotherapy. *J. Clin. Invest.* **102**, 98–106 (1998).
- Jutel, M. *et al.* Bee venom immunotherapy results in decrease of IL-4 and IL-5 and increase of IFN- γ secretion in specific allergen-stimulated T-cell cultures. *J. Immunol.* **154**, 4187–4194 (1995).
- Faith, A., Akdis, C. A., Akdis, M., Simon, H. U. & Blaser, K. Defective TCR stimulation in anergized Th2 cells correlates with abrogated p56lck and ZAP70 tyrosine kinase activities. *J. Immunol.* **159**, 53–60 (1997).
- Watanabe, N., Katakura, K., Kobayashi, A., Okumura, K. & Ovary, Z. Protective immunity and eosinophilia in IgE-deficient SJA/9 mice infected with *Nippostrongylus brasiliensis* and *Trishinella spiralis*. *Proc. Natl Acad. Sci. USA* **85**, 4460–4462 (1988).

Acknowledgements

We thank R. M. Zinkernagel, F. Melchers and J. E. DeVries for critically reviewing the manuscript, as well as C. H. Heusser and S. Alkan for anti-IL-4 and anti-IFN- γ antibodies. This work was sponsored by the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to M.J. (e-mail: marek.jutel@siaf.unizh.ch) or C.A.A. (e-mail: akdisac@siaf.unizh.ch).

Auxin transport inhibitors block PIN1 cycling and vesicle trafficking

Niko Geldner^{*†}, Jiří Friml^{†‡§||}, York-Dieter Stierhof^{*}, Gerd Jürgens^{*} & Klaus Palme[‡]

^{*} Zentrum für Molekularbiologie der Pflanzen, Universität Tübingen, Auf der Morgenstelle 3, D-72076 Tübingen, Germany

[‡] Max-Delbrück-Laboratorium in der Max-Planck-Gesellschaft, D-50829 Köln, Germany

[§] Department of Biochemistry, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic

^{||} These authors contributed equally to this work

Polar transport of the phytohormone auxin mediates various processes in plant growth and development, such as apical dominance, tropisms, vascular patterning and axis formation^{1,2}. This view is based largely on the effects of polar auxin transport inhibitors. These compounds disrupt auxin efflux from the cell but their mode of action is unknown³. It is thought that polar auxin flux is caused by the asymmetric distribution of efflux carriers acting at the plasma membrane⁴. The polar localization of efflux carrier candidate PIN1 supports this model⁴. Here we show that the seemingly static localization of PIN1 results from rapid actin-dependent cycling between the plasma membrane and endosomal compartments. Auxin transport inhibitors block PIN1 cycling and inhibit trafficking of membrane proteins that are unrelated to auxin transport. Our data suggest that PIN1 cycling is of central importance for auxin transport and that auxin transport inhibitors affect efflux by generally interfering with membrane-trafficking processes. In support of our conclusion, the vesicle-trafficking inhibitor brefeldin A mimics physiological effects of auxin transport inhibitors.

Polar auxin transport inhibitors are important tools in assessing the role of polar auxin transport in plant development^{5–7}. They disrupt auxin efflux from the cell without interacting with the efflux carrier itself^{8,9}. However, their mechanism of action has remained elusive despite considerable efforts towards the biochemical characterization of binding sites^{10–12} and several screens for auxin-transport-inhibitor mutants^{13–16}. *In situ* visualization of *Arabidopsis* putative auxin efflux carriers PIN1 and PIN2 correlated polar auxin transport with their asymmetric distribution in the plasma membrane^{17,18}. This coordinated polar localization of PIN1 is not established in *gnom* mutant embryos. *GNOM* encodes a brefeldin A (BFA)-sensitive regulator of vesicle trafficking¹⁹. Consistent with this, the vesicle-trafficking inhibitor BFA affects auxin efflux from the cell^{8,9}, abolishes polar localization of PIN1 and induces its intracellular accumulation¹⁹.

BFA-induced PIN1 accumulation in two large intracellular compartments was completely and rapidly reversible (Fig. 1a–c). To determine whether this reversible BFA effect reflects protein syn-

thesis and degradation or continuous cycling between the plasma membrane and endosomal compartments, we inhibited protein synthesis by cycloheximide (CHX). Incubation of roots in 50 μ M CHX for 30 min reduced ³⁵S-labelled methionine incorporation into proteins to below 10% of the control value (data not shown). However, treatment with 50 μ M CHX for 4 h had no detectable effect on the amount of labelled PIN1 at the plasma membrane (Fig. 1d), suggesting that PIN1 protein is turned over slowly. CHX did not interfere with the reversible BFA effect as PIN1 still accumulated in endomembrane compartments (Fig. 1e) and, on withdrawal of BFA, reappeared at the plasma membrane (Fig. 1f). Thus, BFA-induced intracellular accumulation of PIN1 resulted from blocking exocytosis of a steady-state pool of PIN1 that rapidly cycles between the plasma membrane and some endosomal compartment.

In animal cells, BFA alters structure and function of endomembrane compartments, especially the Golgi apparatus, which fuses with other endomembranes^{20–22}. BFA seems to have different effects in plant cells. In BFA-treated maize root cells, Golgi stack systems progressively disappear and aggregates of vesicles accumulate in their vicinity, with Golgi markers localizing to the newly formed

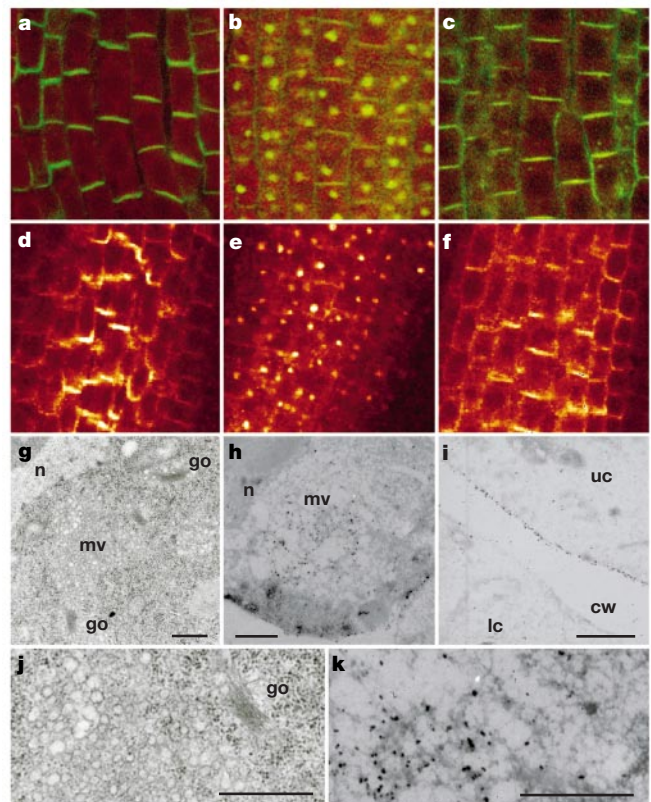


Figure 1 Reversible inhibition of PIN1 recycling by BFA treatment. **a–f**, *Arabidopsis* seedling roots treated with BFA or cycloheximide (CHX) as indicated. **a**, Two-hour buffer control. **b**, Treatment with 50 μ M BFA for 2 h. **c**, Treatment with 50 μ M BFA for 2 h followed by 2 h washing out. **d**, Treatment with 50 μ M CHX for 2 h. **e**, Pre-treatment with 50 μ M CHX for 30 min, then 50 μ M BFA and 50 μ M CHX for 90 min. **f**, Treatment with 50 μ M BFA for 90 min, then 50 μ M BFA and 50 μ M CHX for 30 min followed by 2 h washing out with 50 μ M CHX. **g–k**, Electron micrographs of root cells after treatment with 50 μ M BFA for 30 min (**g**, **h**) and untreated control (**i**). **g**, Ultrastructure. **h**, **i**, Anti-PIN1 immunogold labelling. Note aggregate of membrane vesicles (mv) in ribosome-free cytoplasm near nucleus (n) and Golgi structures (go) of treated cells (**g**, **h**). **i**, Untreated control. Note PIN1 label at the plasma membrane of the upper cell (uc) but not the lower cell (lc). cw, cell wall. **j**, Higher magnification of **g**. **k**, Higher magnification of **h**. Scale bars, 1 μ m.

^{||} Present address: Zentrum für Molekularbiologie der Pflanzen, Universität Tübingen, Auf der Morgenstelle 3, D-72076 Tübingen, Germany.

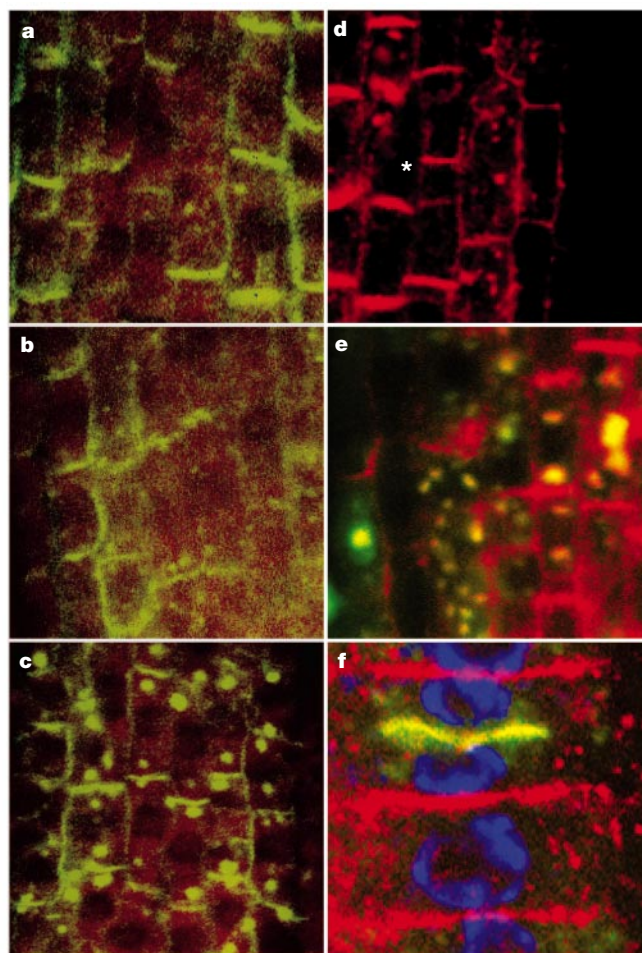


Figure 2 PIN1 localization affected by cytoskeleton-depolymerizing drugs. **a–c**, Cytochalasin D (cytD) effect on BFA inhibition of PIN1 (green) cycling. **a**, Treatment with 20 μM cytD for 2 h. **b**, Pre-treatment with 20 μM cytD for 15 min, then 50 μM BFA and 20 μM cytD for 45 min. **c**, Treatment with 50 μM BFA for 45 min followed by 90 min washing out of BFA with 20 μM cytD. **d**, PIN1 staining after 10 μM oryzalin for 2 h—note patches in cell marked with asterisk as compared with normal localization in cell below. **e, f**, Colocalization of PIN1 (red) and KNOLLE (green) in small patches (yellow) after treatment with 10 μM oryzalin for 2 h (**e**), and in plane of cell division (yellow) of untreated control (**f**). DAPI staining of nuclei is blue.

BFA compartments^{23,24}. In *Arabidopsis* root cells, the BFA-induced PIN1-labelled compartments were juxtanuclear, of a specific size and shape, and mostly two per cell (Fig. 1b). By electron microscopy, we identified aggregates of membrane vesicles of the corresponding size and shape that appeared only after BFA treatment (Fig. 1g, j). The vesicle aggregates were apparent within 30 min of BFA treatment, conditions that left individual Golgi stacks almost intact (Fig. 1g, j). These vesicle aggregates contained PIN1 protein that was apparently internalized from the plasma membrane (Fig. 1h, i, k). This result suggests that BFA compartments consist of vesicle aggregates that derive not only from Golgi stacks, as shown previously^{23,24}, but also from an as yet unidentified endosomal compartment.

The trafficking of membrane vesicles to sites of delivery takes place along the cytoskeleton^{21,22}. To determine cytoskeletal requirements of PIN1 localization, we treated seedling roots with cytochalasin D (cytD) or latrunculin B, which both depolymerize actin filaments^{25,26}, or with oryzalin, which depolymerizes microtubules²⁵. Cytochalasin D alone had only a slight effect on PIN1 polar accumulation at the plasma membrane, which was reduced, with

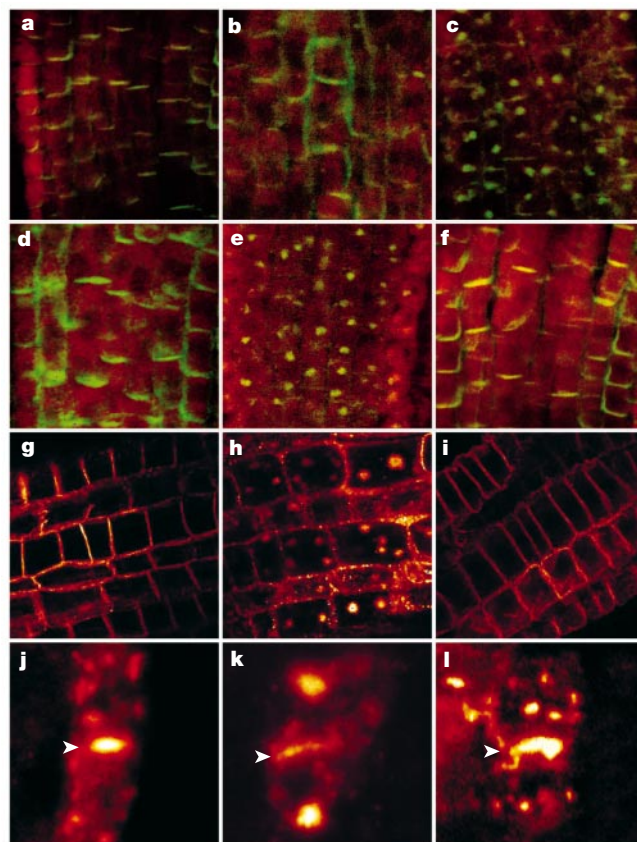


Figure 3 Effects of polar auxin transport inhibitor TIBA on protein trafficking. **a–f**, PIN1. **a**, Treatment with 25 μM TIBA for 2 h. **b**, Pre-treatment with 25 μM TIBA for 30 min, then 50 μM BFA and 25 μM TIBA for 2 h. **c**, Treatment with 50 μM BFA for 2 h followed by 2 h washing out of BFA with 25 μM TIBA. **d**, Treatment with 50 μM benzoic acid for 2 h (compare with **a**). **e**, Pre-treatment with 50 μM benzoic acid for 30 min, then 50 μM BFA and 50 μM benzoic acid for 2 h (compare with **b**). **f**, Treatment with 50 μM BFA for 2 h followed by 2 h washing out of BFA with 50 μM benzoic acid (compare with **c**). Note that BFA inhibition of PIN1 cycling was blocked only by TIBA but not by benzoic acid, which has no inhibitory activity on auxin transport. **g–i**, PM-ATPase. **g**, Untreated control. **h**, Treatment with 50 μM BFA for 90 min. **i**, Pre-treatment with 50 μM TIBA for 30 min, then 50 μM BFA and 50 μM TIBA for 90 min. **j–l**, KNOLLE. **j**, Untreated control. **k**, Treatment with 50 μM BFA for 2 h. **l**, Pre-treatment with 50 μM TIBA for 30 min, then 50 μM BFA and 50 μM TIBA for 2 h. Arrowheads mark cell plates. Note aggregation of KNOLLE label in large patches in **k** as compared with **j** and **l**.

small irregular dots of PIN1 label sometimes appearing underneath the plasma membrane (Fig. 2a). By contrast, cytD inhibited BFA-induced intracellular PIN1 accumulation (Fig. 2b) as well as its relocalization to the plasma membrane when BFA was washed out (Fig. 2c; compare with Fig. 1b, c). Latrunculin B had the same effects (data not shown). These results suggest that PIN1 cycling is actin dependent. By contrast, oryzalin treatment for up to 6 h did not alter PIN1 localization (Fig. 2d) nor internalization in response to BFA (data not shown), although the microtubules were completely depolymerized (data not shown). However, oryzalin treatment caused irregular patches of labelled PIN1 in some cells (Fig. 2d) that were shown to always coexpress the cytokinesis-specific syntaxin KNOLLE (Fig. 2e), which localizes to Golgi stacks and cell plate^{27,28}. PIN1 and KNOLLE colocalized in oryzalin-treated cells (Fig. 2e), and both proteins were also detected at the cell plate of untreated cells (Fig. 2f). Thus, PIN1 seems to traffic along two different pathways, depending on the phase of the cell cycle: an actin-dependent interphase pathway to and from the plasma membrane and a microtubule-dependent cytokinesis pathway to the cell plate. PIN1 accumulation at the newly formed cell plate

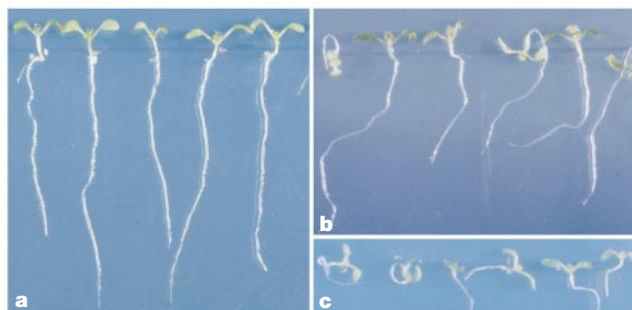


Figure 4 Physiological effects of BFA. Seedlings germinated on vertical agar plates containing different concentrations of BFA. **a**, Untreated control. **b**, Treatment with 10 μ M

BFA. **c**, Treatment with 20 μ M BFA. Note the effect on gravitropic root growth and root elongation in **b** and **c**.

raises the question of how the coordinated polar localization of PIN1 is re-established after cell division, as only one of the two daughter cells would accumulate PIN1 at the correct end. The other daughter cell would initially have PIN1 localized at both ends, necessitating withdrawal of PIN1 from one end to reconstitute the original polarity of the cell.

Polar auxin transport inhibitors, such as 2,3,5-triiodobenzoic acid (TIBA), had only slight effects on PIN1 localization on their own (Fig. 3a). However, clear-cut effects were observed in BFA-treated cells. When cells were pre- and co-treated with TIBA, BFA did not induce intracellular accumulation of PIN1, which instead remained at the plasma membrane (Fig. 3b). Conversely, PIN1 remained at an intracellular position when BFA was washed out in the presence of TIBA (Fig. 3c). By contrast, an inactive analogue of TIBA, benzoic acid, did not interfere with the BFA-induced reversible internalization of PIN1 (Fig. 3d–f). Thus, BFA treatment revealed that polar auxin transport inhibitors block both the removal of PIN1 from and its redistribution to the plasma membrane.

To determine whether polar auxin transport inhibitors affect PIN1 cycling specifically, we analysed their effects on the localization of other membrane proteins whose trafficking is sensitive to BFA. Plasma membrane H^+ -ATPase (PM-ATPase) is turned over rapidly at the plasma membrane²⁹. Indeed, BFA treatment led to reversible intracellular accumulation of PM-ATPase in epidermal cells just above the cell division zone of the root tip. The BFA-induced intracellular compartments of PM-ATPase accumulation were similar in size and shape to those that accumulated PIN1 (Fig. 3g, h). In the presence of TIBA, BFA did not induce intracellular accumulation of PM-ATPase (Fig. 3i). Furthermore, recovery from the effect of BFA was blocked by washing with TIBA (data not shown). Thus, PM-ATPase trafficking was sensitive to polar auxin transport inhibitors, as was PIN1 trafficking. We also tested KNOLLE syntaxin²⁷. With BFA treatment, small dots of labelled KNOLLE aggregated into larger patches (Fig. 3j, k). This BFA effect was also abolished by co-treatment with TIBA (Fig. 3l). Every polar auxin transport inhibitor that we tested had similar effects on the localization of all three proteins (PIN1, KNOLLE and PM-ATPase), although the effective concentrations were different, with 25 μ M TIBA or 1-pyrenoylbenzoic acid (PBA) leading to complete inhibi-

tion of BFA-visualized trafficking, as compared with approximately 200 μ M for 9-hydroxyfluorene-9-carboxylic acid (HFCA) or 1-N-naphthylphthalamic acid (NPA). Thus, a common feature of auxin-transport-inhibitor action seems to be a rather general influence on protein trafficking. We therefore tested whether the well established vesicle-trafficking inhibitor BFA might mimic physiological effects of auxin transport inhibitors. *Arabidopsis* seedlings were grown on vertical agar plates supplemented with low concentrations of BFA (Fig. 4). Several processes known to be sensitive to auxin transport inhibitors, such as root and hypocotyl gravitropism and elongation as well as lateral root initiation, were affected by BFA in a concentration-dependent manner (Fig. 4; Table 1). It is conceivable that compounds that were originally identified as polar auxin transport inhibitors may actually be inhibitors of membrane trafficking.

Our observations indicate that the putative auxin-efflux carrier PIN1 cycles between the plasma membrane and endosomal compartments in an actin-dependent manner. Auxin transport inhibitors block PIN1 cycling but, unlike BFA, leave PIN1 accumulation at the plasma membrane unaffected. This implies that cycling of some component of the auxin efflux machinery is essential for the mechanism of polar auxin transport. Notably, polar auxin transport inhibitors seem to interfere with membrane trafficking in broader terms, suggesting that their seemingly specific physiological effects may reflect a high sensitivity of auxin efflux carrier cycling and the important role of polar auxin flow in plant growth and development. □

Methods

Plant material and growth conditions

We grew *Arabidopsis* seedlings (Columbia ecotype) on vertical agar plates at 25 °C for 7 days.

Inhibitor treatments

Ten–fifteen seedlings were incubated in 1 ml liquid growth medium (0.5×MS medium, 0.8% sucrose, pH 5.8) containing the respective inhibitors or equal amounts of solvents (controls) in 24-well cell-culture plates. Washing was done twice for 10 min and a third washing step was carried out up to the indicated time. Incubations were stopped by fixation. BFA (Molecular Probes), TIBA (Duchefa), NPA (Duchefa), HFCA (Sigma) and PBA (synthesized according to ref. 30) were used as 100 mM stock in DMSO, cytD (Sigma) and oryzalin (Duchefa) as 100 mM stock in ethanol, and cycloheximide (Sigma) as 50 mM stock in water.

Whole-mount immunofluorescence analysis on root tips

Immunofluorescence analysis was carried out as described²⁷ or according to ref. 18 using an In SituPro machine (Intavis) for changing solutions. Rabbit polyclonal anti-PIN1 (ref. 17) was diluted at 1:200, mouse monoclonal anti-PM-ATPase at 1:500 and rabbit polyclonal anti-KNOLLE²⁷ at 1:4,000. Double labelling of KNOLLE PIN1 was done with polyclonal rat anti-KNOLLE serum. CY3- and fluorescein-isothiocyanate-conjugated secondary antibodies were from Dianova. We carried out confocal laser scanning microscopy with Leica microscopes.

Table 1 Quantitative effects of different BFA concentrations

Area affected	Concentration of BFA			
	0 μ M	1 μ M	10 μ M	20 μ M
Number of lateral roots	3.05 $\pm$ 1.21	3.48 $\pm$ 0.94	0.35 $\pm$ 0.49	0.10 $\pm$ 0.31
Root length (mm)	39.87 $\pm$ 3.66	37.70 $\pm$ 5.12	23.47 $\pm$ 4.91	7.07 $\pm$ 1.79
Hypocotyl length (mm)	23.1 $\pm$ 2.2	22.2 $\pm$ 2.4	18.6 $\pm$ 2.0	9.9 $\pm$ 2.7

Immunolocalization on cryosections and electron microscopy

Ultrastructure analysis of chemically fixed sections and immunogold labelling using silver-enhanced nanogold of ultrathin cryosections were carried out as described²⁹.

Received 3 July; accepted 9 August 2001.

1. Estelle, M. Polar auxin transport. New support for an old model. *Plant Cell* **10**, 1775–1778 (1998).
2. Berleth, T. & Sachs, T. Plant morphogenesis: long-distance coordination and local patterning. *Curr. Opin. Plant Biol.* **4**, 57–62 (2001).
3. Morris, D. A. Transmembrane auxin carrier systems—dynamic regulators of polar auxin transport. *Plant Growth Reg.* **32**, 161–172 (2000).
4. Palme, K. & Gälweiler, L. PIN-pointing the molecular basis of auxin transport. *Curr. Opin. Plant Biol.* **2**, 375–381 (1999).
5. Hadfi, K., Speth, V. & Neuhaus, G. Auxin-induced developmental patterns in *Brassica juncea* embryos. *Development* **125**, 879–887 (1998).
6. Mattsson, J., Sung, Z. R. & Berleth, T. Responses of plant vascular systems to auxin transport inhibition. *Development* **126**, 2979–2691 (1999).
7. Sabatini, S. *et al.* An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99**, 463–472 (1999).
8. Delbarre, A., Muller, P. & Guern, J. Short-lived and phosphorylated proteins contribute to carrier-mediated efflux, but not to influx, of auxin in suspension-cultured tobacco cells. *Plant Physiol.* **116**, 833–844 (1998).
9. Morris, D. A. & Robinson, J. S. Targeting of auxin carriers to the plasma membrane: differential effects of brefeldin A on the traffic of auxin uptake and efflux carriers. *Planta* **205**, 606–612 (1998).
10. Michalke, W., Katekar, G. F. & Geissler, A. E. Phytotropin-binding sites and auxin transport in *Cucurbita pepo*: evidence for two recognition sites. *Planta* **187**, 254–260 (1992).
11. Cox, D. N. & Muday, G. K. NPA binding activity is peripheral to the plasma membrane and is associated with the cytoskeleton. *Plant Cell* **6**, 1941–1953 (1994).
12. Bernasconi, P., Patel, B. C., Reagan, J. D. & Subramanian, M. V. The N-1-naphthylphthalamic acid-binding protein is an integral membrane protein. *Plant Physiol.* **111**, 427–432 (1996).
13. Fujita, H. & Syono, K. Genetic analysis of the effects of polar auxin transport inhibitors on root growth in *Arabidopsis thaliana*. *Plant Cell Physiol.* **37**, 1094–1101 (1996).
14. Garbers, C. *et al.* A mutation in protein phosphatase 2A regulatory subunit A affects auxin transport in *Arabidopsis*. *EMBO J.* **15**, 2115–2124 (1996).
15. Ruegger, M. *et al.* Reduced naphthylphthalamic acid binding in the *tir3* mutant of *Arabidopsis* is associated with a reduction in polar auxin transport and diverse morphological defects. *Plant Cell* **9**, 745–757 (1997).
16. Fujita, H. & Syono, K. *PIS1*, a negative regulator of the action of auxin transport inhibitors in *Arabidopsis thaliana*. *Plant J.* **12**, 583–595 (1997).
17. Gälweiler, L. *et al.* Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. *Science* **282**, 2226–2230 (1998).
18. Müller, A. *et al.* AtPIN2 defines a locus of *Arabidopsis* for root gravitropism control. *EMBO J.* **17**, 6903–6911 (1998).
19. Steinmann, T. *et al.* Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science* **286**, 316–318 (1999).
20. Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. Brefeldin A: insights into the control of membrane traffic and organelle structure. *J. Cell Biol.* **116**, 1071–1080 (1992).
21. Lippincott-Schwartz, J. *et al.* Microtubule-dependent retrograde transport of proteins into the ER in the presence of brefeldin A suggests an ER recycling pathway. *Cell* **60**, 821–836 (1990).
22. Wood, S. A., Park, J. E. & Brown, W. J. Brefeldin A causes a microtubule-mediated fusion of the trans-Golgi network and early endosomes. *Cell* **67**, 591–600 (1991).
23. Satiat-Jeunemaitre, B. & Hawes, C. Redistribution of a Golgi glycoprotein in plant cells treated with brefeldin A. *J. Cell Sci.* **103**, 1153–1156 (1992).
24. Satiat-Jeunemaitre, B. *et al.* Brefeldin A effects in plant and fungal cells: something new about vesicle trafficking? *J. Microsc.* **181**, 162–177 (1996).
25. Satiat-Jeunemaitre, B., Steele, C. & Hawes, C. Golgi-membrane dynamics are cytoskeleton dependent: a study on Golgi stack movement induced by brefeldin A. *Protoplasma* **191**, 21–33 (1996).
26. Mathur, J., Spielhofer, P., Kost, B. & Chua, N.-H. The actin cytoskeleton is required to elaborate and maintain spatial patterning during trichome cell morphogenesis in *Arabidopsis thaliana*. *Development* **126**, 5559–5568 (1999).
27. Lauber, M. H. *et al.* The *Arabidopsis* KNOLLE protein is a cytokinesis-specific syntaxin. *J. Cell Biol.* **139**, 1485–1493 (1997).
28. Völker, A., Stierhof, Y.-D. & Jürgens, G. Cell cycle-independent expression of the *Arabidopsis* cytokinesis-specific syntaxin KNOLLE results in mistargeting to the plasma membrane and is not sufficient for cytokinesis. *J. Cell Sci.* (in the press).
29. Hager, A. *et al.* Auxin induces exocytosis and the rapid synthesis of a high-turnover pool of plasma-membrane H⁺-ATPase. *Planta* **185**, 527–537 (1991).
30. Kategar, G. F. & Geissler, A. E. Auxin transport inhibitors. *Plant Physiol.* **60**, 826–829 (1977).

Acknowledgements

We thank A. Vieten and S. Barth for technical assistance; L. Gälweiler, W. Michalke and K. Roberts for kind gifts of antibodies; S. M. Li for MS/NMR analysis; and M. Godde, M. Heese, T. Hamann, T. Pacher and K. Schrick for helpful comments and critical reading of the manuscript. J. F. was a recipient of a *Deutscher Akademischer Austauschdienst* fellowship. This work was supported by Deutsche Forschungsgemeinschaft (DFG), Schwerpunktprogramm 'Molekulare Analyse der Phytohormonwirkung', European Communities Biotechnology Programme and INCO Copernicus Programme (K.P.), and DFG, Sonderforschungsbereich 'Mechanismen des Zellverhaltens bei Eukaryoten' (G. J.).

Correspondence and requests for materials should be addressed to G.J. (e-mail: gerd.juergens@zmbp.uni-tuebingen.de).

Voltage-induced membrane movement

Ping-Cheng Zhang, Asbed M. Keleshian & Frederick Sachs

HHMI Center for Single Molecule Biophysics, State University of New York at Buffalo, Buffalo, New York 14214, USA

Thermodynamics predicts that transmembrane voltage modulates membrane tension¹ and that this will cause movement. The magnitude and polarity of movement is governed by cell stiffness and surface potentials. Here we confirm these predictions using the atomic force microscope to dynamically follow the movement of voltage-clamped HEK293 cells² in different ionic-strength solutions. In normal saline, depolarization caused an outward movement, and at low ionic strength an inward movement. The amplitude was proportional to voltage (about 1 nm per 100 mV) and increased with indentation depth. A simple physical model of the membrane and tip provided an estimate of the external and internal surface charge densities ($-5 \times 10^{-3} \text{ C m}^{-2}$ and $-18 \times 10^{-3} \text{ C m}^{-2}$, respectively). Salicylate (a negative amphiphile³) inhibited electromotility by increasing the external charge density by $-15 \times 10^{-3} \text{ C m}^{-2}$. As salicylate blocks electromotility in cochlear outer hair cells at the same concentration^{4,5}, the role of prestin as a motor protein⁶ may need to be reassessed.

The Lippman¹ equation relates interfacial tension to electrostatic potential. The physical basis for this thermodynamic relationship stems from properties of the double layer adjacent to polarizable interfaces. This region contains excess mobile charges that are attracted to the fixed structural charges or are supplied by external voltage sources. These mobile charges repel each other laterally, creating a local pressure and changing the net surface tension. In lipid monolayers, changes in ionic strength alter the surface pressure (tension)⁷. Membrane movements correlated with action potentials have been observed in crab nerve as well as squid giant axon⁸. In a bilayer membrane with two interfaces, a difference in tension between the interfaces can cause changes in curvature, an effect also known as converse flexoelectricity. Both flexoelectricity (current generation from bending) and its converse have been demonstrated in lipid bilayers⁹ and cell membranes¹⁰.

When an atomic force microscope (AFM) tip indents a membrane, changes in membrane tension will move the cantilever². We have measured these displacements in real time using voltage-clamped HEK293 cells (Fig. 1a). A simple model that applies the Lippmann equation to both interfaces explains the observed movement over a wide range of surface potentials and applied voltage. As the Lippmann equation is derived from thermodynamic principles, the basic model should apply to all membranes.

As can be seen from Fig. 1b, in normal saline hyperpolarizing pulses cause the cantilever to move towards the cell interior, whereas depolarization reverses the movement. At low ionic strength (isotonic 288 mM sucrose, ionic strength 20 μM measured by conductivity) the potential dependence inverts. The time-dependent changes seen in Fig. 1b may result from cytoskeletal relaxation or lipid flip-flop (the movement of lipid molecules between the internal and external monolayers), and are the subject of an ongoing study. The peak displacement is linear with voltage (Fig. 1c) and increases with the mean indentation force (Fig. 1d). The peak displacement as a function of the bath ionic strength is shown in Fig. 2a for both hyperpolarizing and depolarizing pulses ($\pm 100 \text{ mV}$, from a holding potential of -60 mV). Note the reversal of movement at about 10 mM.

We modelled the system by applying the Lippmann equation at both interfaces (assumed independent¹¹) and treating the bilayer as an incompressible capacitor. Whereas the mean force on an indent-

ing AFM tip depends primarily on the cytoplasmic compliance (Fig. 2d), the voltage-dependent modulation arises from global alterations in membrane tension (Fig. 2c). A useful analogy is the way in which a weight (cantilever tip) on a clothesline (membrane) is moved vertically by changes in the tension. The Lippmann tension depends on the concentration of mobile charges in the double layer, which in turn is a function of the ionic strength of the solution bathing the interface, the fixed charge density and the transmembrane potential. The Lippmann equation for a single interface is: $d\gamma/d\Psi_D = -q$, where γ is the surface tension, Ψ_D is the surface potential and q is the excess mobile charge in the double layer. By approximating the double layer as a parallel plate capacitor¹, the integral of the Lippmann equation yields: $\gamma = -0.5C_D\Psi_D^2 + T_0$, where C_D is the specific capacitance of the double layer, and T_0 is the voltage-independent tension. The sign convention reflects that increasing charge stretches the interface. C_D is calculated by the parallel plate model of Helmholtz with plate spacing equal to the Debye length (κ_D) and a relative permittivity for water (ϵ_w) in the range of 40–80. For solutions with equal valence ions (z), $\kappa_D = (\epsilon_w\epsilon_0k_BT/2z^2e_0^2n)^{1/2}$ and $C_D = \epsilon_w\epsilon_0/\kappa_D$ (where e_0 is the electronic charge, ϵ_0 the permittivity of free space, k_BT Boltzmann's constant times the absolute temperature, and n the ionic strength of the solution). In physiological saline $\kappa_D \approx 1$ nm and $C_D \approx 30 \mu\text{F cm}^{-2}$. If the structural charges are located only at the interface, the surface potential can be obtained from the Poisson–Boltzmann equation:

$$\Psi_D = \frac{2k_BT}{ze_0} \sinh^{-1} \left(\frac{q}{2\sqrt{n\epsilon_w\epsilon_0}2k_BT} \right)$$

Although this simple model ignores steric factors, hydration layers and adsorption, it has proven surprisingly effective in modelling electrical properties of the double layer¹².

With the above assumptions and considering the two interfaces as independent, the membrane is approximated as three capacitors in series: a Debye capacitance for each interface and a hydrocarbon layer in between (with a specific capacitance of $C_m \approx 0.5 \mu\text{F cm}^{-2}$). The Lippmann tension is only a weak function of the applied potential as $C_D \gg C_m$, and most of the applied voltage (approximately 98%) drops across C_m . Hence, the voltage-induced mobile charge accumulation at either interface can also be approximated as $C_m V$ (where V is the applied potential). The total tension is the sum of the two interface tensions. For a sense of scale, the lytic tension for biological membranes is in the order of 0.01 N m^{-1} (10 dyn cm^{-1})¹³ and even at an applied voltage of 1 V (in a neutral interface) the Lippmann tension will be less than 1% of the lytic limit. The total tension in the membrane (T_t) is then:

$$T_t = \frac{\sqrt{(2k_BT)^3\epsilon_w\epsilon_0}}{ze_0} \left(\sqrt{n_{\text{ex}}} \left[\sinh^{-1} \left\{ \frac{\sigma_{\text{ex}} - C_m V}{2\sqrt{n_{\text{ex}}\epsilon_w\epsilon_0}2k_BT} \right\} \right]^2 + \sqrt{n_{\text{in}}} \left[\sinh^{-1} \left\{ \frac{\sigma_{\text{in}} + C_m V}{2\sqrt{n_{\text{in}}\epsilon_w\epsilon_0}2k_BT} \right\} \right]^2 \right) + T_m$$

where T_m is the voltage-independent portion of the total membrane tension (equivalent to the lytic tension in this formulation), σ is the structural charge density at the interface and the subscripts 'ex' and 'in' refer to the external and internal interfaces respectively.

T_t produces a normal force proportional to the length over which it acts. For an AFM tip approximated as a 70° cone, this length (L) is equal to the contact perimeter where the membrane separates tangentially from the tip. The relationship between the mean force and the contact perimeter is given by $F = EL^2 \cot(35^\circ) / (8\pi(1 - \nu^2))$ from hertzian mechanics¹⁴ with E as Young's modulus and ν as Poisson's ratio. The normal component of $F = T_t \cos(35^\circ)L$ (Fig. 2d). The expression can be simplified by defining an 'effective

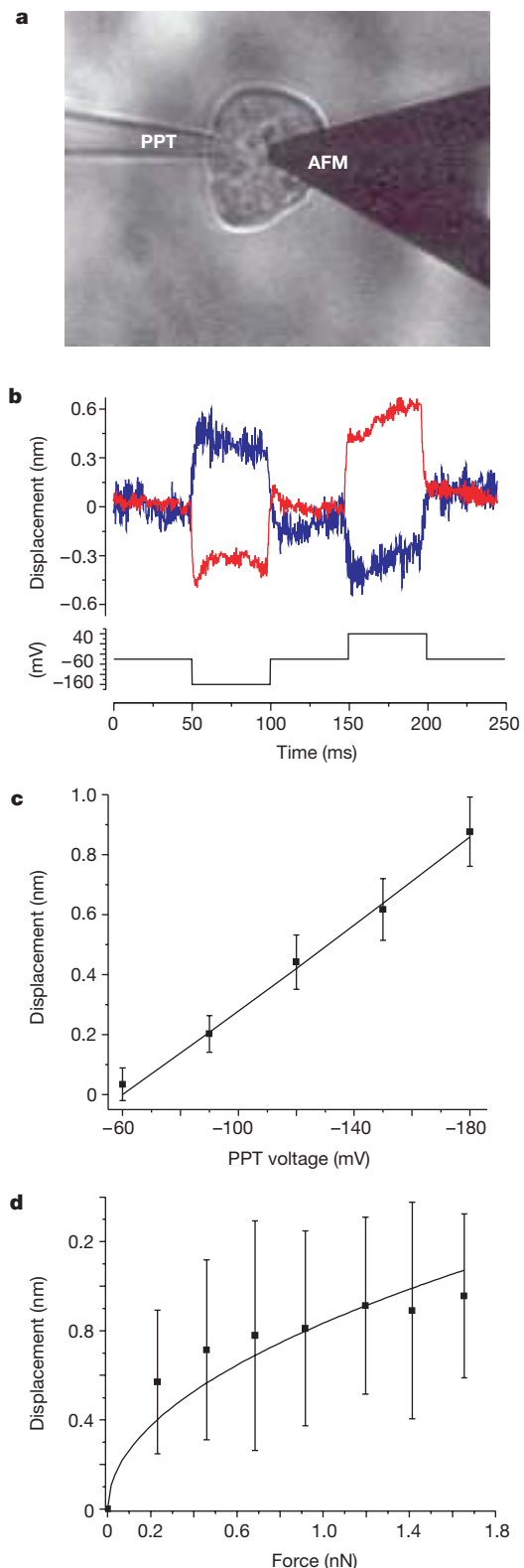


Figure 1 Voltage-induced membrane movement in HEK293 cells. **a**, Micrograph of a voltage-clamped cell with the cantilever (AFM) in place. The pipette (PPT) can be seen on the left side. **b**, Representative movements driven by ± 100 -mV voltage pulses from a holding potential of -60 mV at normal (blue trace) and nominal 0 mM (red trace) ionic strengths. A positive displacement corresponds to the cantilever moving into the cell. **c**, The movement is linear with voltage (normal ionic strength). Solid line is the fit from the model (see text). **d**, Movement increases with indentation force. Solid line is the fit from the model, assuming hertzian mechanics (see text). Error bars represent the s.d.

path length' $L_{\text{eff}} = L \cos(35^\circ)$, such that $T L_{\text{eff}}$ then gives the force acting on the cantilever. A change in tension ΔT_t acting along L_{eff} modulates the mean force F by ΔF . The voltage-induced cantilever movement Δz is obtained from the spring constant of the cantilever ($k_{\text{AFM}} = 0.01 \text{ N m}^{-1}$): $\Delta z = \Delta F/k_{\text{AFM}} = \Delta T_t L_{\text{eff}}/k_{\text{AFM}}$. Note that a higher mean indentation force would increase L_{eff} and produce a larger AFM tip displacement for a given voltage change.

Our data suggest that at low ionic strength the ionic efflux from the cell would increase the local ionic strength, causing significant deviations from the bath concentration. Diffusion predicts a minimum accumulation of approximately $160 \mu\text{M nA}^{-1}$ for a $10\text{-}\mu\text{m}$ diameter spherical cell¹⁵. Consequently we added to the model a parameter for accumulation (n_{acc}) at the extracellular surface.

Fitting the model to the data, we can extract four parameters: the two surface charge densities, the perimeter of the contact region and the mean ionic accumulation. The solid lines of Fig. 2a correspond to the optimal fit (assuming $n_{\text{in}} = 0.17 \text{ mM}$)¹⁶. The mean external and internal surface potentials (in normal ionic-strength bath solution) were -7 and -25 mV , respectively. Although the external potential estimated from the zeta potentials of primary kidney cells was -30 mV (ref. 17), the discrepancy can be explained from the use of different methodologies. Zeta potentials include all charges inside the hydrodynamic shear plane¹⁸, such as negatively charged glycoproteins, whereas our measurements are dominated by charges close to the hydrocarbons. The inner surface potential is significantly more negative than the external. Although fluorescent probe studies have measured surface potentials in bilayers¹⁹, we were unable to find data for cell membranes. The surface charges may arise from proteins or lipids and although measurements on ion channels

suggest a more negative exterior than interior¹⁶, those measurements are local to the channels.

The accumulation parameter n_{acc} was 0.42 mM nA^{-1} (larger than the predicted minimum of 0.16 mM nA^{-1}), but it was much less than expected in the restricted volume between the membrane and

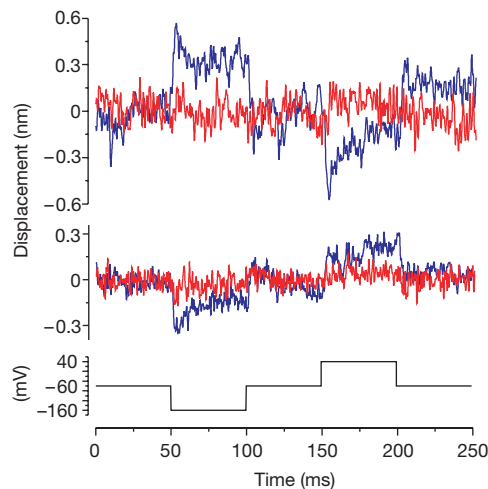


Figure 3 Reduction of voltage-induced membrane displacement in the presence of salicylate. Salicylate (4 mM) reduced electromotility in HEK cells in both normal ionic-strength (top; 184 mM) and low ionic-strength (bottom; 2 mM) bath solutions. The blue traces are controls; red traces are in the presence of sodium salicylate (traces are from different cells). The effect was reversible on washout (traces not shown).

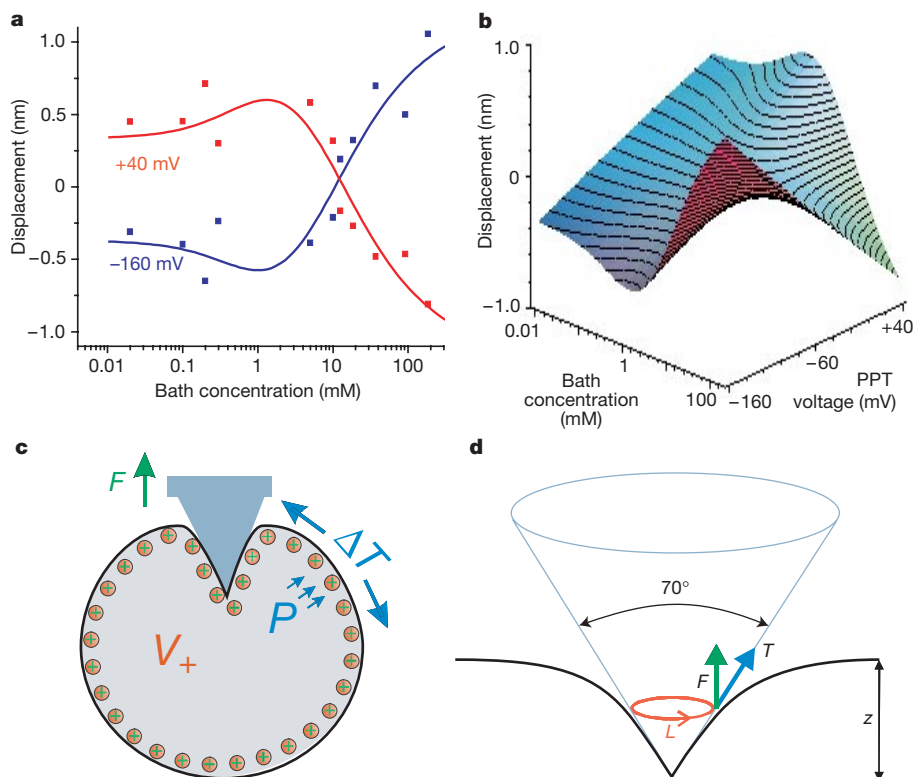


Figure 2 Experimental and predicted voltage-induced AFM movement in differing ionic-strength solutions. **a**, AFM displacement versus ionic strength. Red symbols represent depolarization; blue hyperpolarization. Each data point is from one or more cells (s.d. of 0.3 nm). The solid lines are the optimal fits from the model described in the text. The estimated parameters are: $\sigma_{\text{ex}} = -5 \pm 1 \text{ mC m}^{-2}$, $\sigma_{\text{in}} = -18 \pm 2 \text{ mC m}^{-2}$, $L_{\text{eff}} = 1.1 \mu\text{m}$, $n_{\text{acc}} = 0.42 \text{ mM}$. **b**, Three-dimensional plot of predicted displacement with respect to

voltage and ionic strength using the estimated parameters. **c**, Lateral repulsion of charges changes surface tension. The equivalent change in pressure (P) to a positive potential (V_+) produces a force displacing the cantilever. **d**, The local geometry of the indentation determines the path length (L) along which the membrane tension (T) acts to produce a normal force (F).

the AFM tip (calculations not shown). Thus, the force-generating mechanism seems global rather than local to the tip. With a mean force of 0.46 nN and a bending modulus of 10^{-19} J the local curvature of a bilayer will match that of the AFM tip (~ 50 nm).

The Lippmann model explains our fundamental experimental observations: a 'linear' relation between Δz and voltage at normal ionic strength (Fig. 1c), an increase in movement at higher indentation forces (Fig. 1d), a sign reversal at ~ 10 mM and a loss of sensitivity to changes in ionic strength at low ionic strength (Fig. 2a, b). The voltage-dependent changes in tension should also be measurable in lipid tethers using optical or magnetic traps²⁰.

Alterations in surface charge by any means should alter the voltage-induced movement. We tested extracellular pH and sodium salicylate (at 4 mM)—a negative amphiphile reported to partition in membrane lipids²¹ and affect surface charge³. We were unable to test pH levels lower than 5.4 owing to seal breakdown, but we found little effect in normal ionic-strength solution, probably from a lack of titratable charges. However, salicylate reversibly reduced displacement in both normal and low ionic-strength solutions (Fig. 3). The model predicts that inhibition of motility can arise only from an increase in cell stiffness (that is, a decrease in L_{eff}) or a more negative external surface charge. Force–distance measurements did not show any statistically significant effect of salicylate (for Hertz model, $E = 0.97$ and 1.15 kPa with and without salicylate, respectively). The reduction in electromotility must be caused by a change in surface charge. The fitting to the model predicted an increase of the external charge density to -20×10^{-3} C m⁻² and (an unmodified) internal density of -18×10^{-3} C m⁻². The preferential distribution of the ionized form of salicylate to the outer monolayer compared with the inner is expected from a holding potential of -60 mV (based on the Nernst potential). (Aspiration studies have suggested that salicylate decreases membrane stiffness in outer hair cells²², but this discrepancy can be explained by the insensitivity of our method to changes in membrane compliance.)

Although salicylate blocks electrical motility in outer hair cells of the cochlea⁴, presumably by interacting directly with the 'motor protein' prestin^{5,23}, the observation that it can also block electromotility in kidney cells (in the absence of prestin) is surprising as outer hair cells have a unique architecture that has been presumed to be associated with the origin of motility. However, the present observations are compatible with the proposition that flexoelectric effects may be the motor underlying outer hair cell motility²⁴, and raise the question of what could be the function of prestin. Outer hair cell electromotility is always associated with a nonlinear capacitance²⁵ and prestin expression in HEK cells produces a large nonlinear capacitance as well as macroscopic voltage-dependent movements²⁶. As the bilayer capacitance is increased by adding mobile charges²⁷, and our model predicts that increased capacitance increases voltage sensitivity, prestin may function primarily as a dipolar charge carrier rather than an 'area motor'⁶.

The ability to measure surface charge densities may serve as a unique assay for the activity of second messengers that are charged amphiphiles, such as the phosphoinositides and arachidonic acid. Furthermore, as surface charge density affects membrane tension, amphipaths can also modify mechanically sensitive channels and other membrane proteins^{28,29}. Another outcome of our analysis is a new measure of cytoskeletal stiffness as we have an independent measure of the contact perimeter of the AFM tip. From fitting the data (Figs 1c, d and 2a), the contact perimeter at 0.46 nN ranged from 0.9 to 1.1 μm (a contact depth of ~ 0.36 μm). This corresponds to a Young's modulus of 5.0–7.5 kPa (assuming a hertzian model with $\nu = 0.5$). These values are higher than traditional force–distance measurements using a first-order hertzian model that underestimates stiffness at large indentations³⁰.

Membranes move with transmembrane potential as expected from thermodynamics. These movements may be physiologically

significant in highly curved membranes such as those in cones, dendritic spines, mitochondria and chloroplasts, and will generate pressure waves. □

Methods

We performed experiments with whole-cell voltage-clamped HEK293 cells. The pipette was filled with (in mM): 145 KCl, 10 EGTA and 5 HEPES. Normal ionic-strength bath contained (in mM): 167 NaCl, 5.6 KCl, 11 HEPES, 0.6 CaCl₂, 0.6 MgCl₂ at pH 7.4. Bath ionic strength was reduced by dilution with sucrose solution (288 mM) keeping the osmolarity at 320 mosM. The cell membrane was indented using the AFM cantilever with a force of 0.46 nN (unless otherwise indicated). For each experiment, we averaged 50 repetitions of the cantilever displacement associated with 50 ms hyperpolarizing and depolarizing pulses of 100 mV (from a holding potential of -60 mV). The displacement was taken as the average of 5 ms about the peak. Positive displacements represent the AFM moving into the cell.

Received 4 April; accepted 8 August 2001.

- Bockris, J. O. & Reddy, A. K. N. *Modern Electrochemistry: an Introduction to an Interdisciplinary Area* Vol. 2 (Plenum, New York, 1973).
- Mosbacher, J., Langer, M., Horber, J. K. & Sachs, F. Voltage-dependent membrane displacements measured by atomic force microscopy. *J. Gen. Physiol.* **111**, 65–74 (1998).
- McLaughlin, S. Salicylates and phospholipid bilayer membranes. *Nature* **203**, 234–236 (1973).
- Shehata, W. E., Brownell, W. E. & Dieler, R. Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pig cochlea. *Acta Otolaryngol* **111**, 707–718 (1991).
- Takehata, S. & Santos-Sacchi, J. Effects of salicylate and lanthanides on outer hair cell motility and associated gating charge. *J. Neurosci.* **16**, 4881–4889 (1996).
- Iwasa, K. H. & Adachi, M. Force generation in the outer hair cell of the cochlea. *Biophys. J.* **73**, 546–555 (1997).
- Yamaguchi, T., Nishizaki, K. & Itai, S. Molecular interactions between phospholipids and electrolytes in a monolayer of parenteral lipid emulsion. *Colloid Surf. B* **9**, 275–282 (1997).
- Iwasa, K., Tasaki, I. & Gibbons, R. C. Swelling of nerve fibers associated with action potentials. *Science* **210**, 338–339 (1980).
- Todorov, A. T., Petrov, A. G. & Fendler, J. H. Flexoelectricity of charged and dipolar bilayer-lipid membranes studied by stroboscopic interferometry. *Langmuir* **10**, 2344–2350 (1994).
- Petrov, A. G., Miller, B. A., Hristova, K. & Usherwood, P. N. Flexoelectric effects in model and native membranes containing ion channels. *Eur. Biophys. J.* **22**, 289–300 (1993).
- Chou, T., Jaric, M. V. & Siggia, E. D. Electrostatics of lipid bilayer bending. *Biophys. J.* **72**, 2042–2055 (1997).
- Winiski, A. P., McLaughlin, A. C., McDaniel, R. V., Eisenberg, M. & McLaughlin, S. An experimental test of the discreteness-of-charge effect in positive and negative lipid bilayers. *Biochemistry* **25**, 8206–8214 (1986).
- Akinlaja, J. & Sachs, F. The breakdown of cell membranes by electrical and mechanical stress. *Biophys. J.* **75**, 247–254 (1998).
- Wu, H. W. & Moy, V. T. Mechanical properties of L929 cells measured by atomic force microscopy: effects of anticytoskeletal drugs and membrane crosslinking. *Scanning* **20**, 389–397 (1998).
- Carlsaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn (Oxford Univ. Press, London, 1959).
- Hille, B. *Ionic Channels of Excitable Membranes* 2nd edn (Sinauer, Sunderland, 1992).
- Singh, A. K., Kasinath, B. S. & Lewis, E. J. Interaction of polycations with cell surface negative charges of epithelial cells. *Biochim. Biophys. Acta* **1120**, 337–342 (1992).
- Carnie, S. & McLaughlin, S. Large divalent cations and electrostatic potentials adjacent to membranes: a theoretical calculation. *Biophys. J.* **44**, 325–332 (1983).
- Kraayenhof, R., Sterk, G. J. & Wong Fong Sang, H. W. Probing biomembrane interfacial potential and pH profiles with a new type of float-like fluorophores positioned at varying distance from the membrane surface. *Biochemistry* **32**, 10057–10066 (1993).
- Hwang, W. C. & Waugh, R. E. Energy of dissociation of lipid bilayer from the membrane skeleton of red blood cells. *Biophys. J.* **72**, 2669–2678 (1997).
- Gutknecht, J. & Tosteson, D. C. Diffusion of weak acids across lipid bilayer membranes: effects of chemical reactions in the unstirred layers. *Science* **182**, 1258–1261 (1973).
- Lue, A. J.-C. & Brownell, W. E. Salicylate induced changes in outer hair cell lateral wall stiffness. *Hearing Res.* **135**, 163–168 (1999).
- Oliver, D. et al. Intracellular anions as the voltage-sensor of prestin, the outer hair cell motor protein. *Science* **292**, 2340–2343 (2001).
- Raphael, R. M., Popel, A. S. & Brownell, W. E. A membrane bending model of outer hair cell electromotility. *Biophys. J.* **78**, 2844–2862 (2000).
- Santos-Sacchi, J. Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *J. Neurosci.* **11**, 3096–3110 (1991).
- Zheng, J. et al. Prestin is the motor protein of cochlear outer hair cell. *Nature* **405**, 149–155 (2000).
- Wu, M. & Santos-Sacchi, J. Effects of lipophilic ions on outer hair cell membrane capacitance and motility. *J. Memb. Biol.* **166**, 111–118 (1998).
- Markin, V. S. & Martinac, B. Mechanosensitive ion channels as reporters of bilayer expansion: a theoretical model. *Biophys. J.* **60**, 1120–1127 (1991).
- Maingret, F., Fosset, M., Lesage, F., Lazdunski, M. & Honore, E. TRAAK is a mammalian neuronal mechano-gated K⁺ channel. *J. Biol. Chem.* **274**, 1381–1387 (1999).
- Shih, C. W., Schleif, W. S. & Li, J. C. M. Photoelastic and finite element analysis of different size spheres in contact. *J. Mater. Res.* **7**, 1011–1017 (1992).

Acknowledgements

We would like to thank K. Snyder and A. Petrov, A. Boulbitch, R. Raphael, J. Santos-Sacchi, O. Anderson and W. Brownell for encouragement and suggestions, and the US Army Research Office, NIH and the Cell Mechanosensing Project, ICORP, Japan Science and

Technology Corporation for support. This work was also funded in part by the Ralph Hochstetter Medical Research Fund.

Correspondence and requests for materials should be addressed to F.S. (e-mail: sachs@buffalo.edu).

The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita

Tom Vulliamy*, Anna Marrone*, Frederick Goldman†, Andrew Dearlove‡, Monica Bessler§, Philip J. Mason* & Inderjeet Dokal*

* Department of Haematology, Division of Investigative Science, Faculty of Medicine, Imperial College School of Science, Technology and Medicine, Hammersmith Hospital, Duane Road, London W12 0NN, UK

† Department of Pediatrics, The University of Iowa Hospitals and Clinics, Iowa City, Iowa 52242-1083, USA

‡ MRC UK, HGMP Resource Centre, Hinxton Cambridge, CB10 1SB, UK

§ Division of Hematology, Washington University School of Medicine, St. Louis, Missouri 63110, USA

Dyskeratosis congenita is a progressive bone-marrow failure syndrome that is characterized by abnormal skin pigmentation, leukoplakia and nail dystrophy^{1,2}. X-linked, autosomal recessive and autosomal dominant inheritance have been found in different pedigrees. The X-linked form of the disease is due to mutations in the gene *DKC1* in band 2, sub-band 8 of the long arm of the X chromosome (ref. 3). The affected protein, dyskerin, is a nucleolar

protein that is found associated with the H/ACA class of small nucleolar RNAs and is involved in pseudo-uridylation of specific residues of ribosomal RNA⁴. Dyskerin is also associated with telomerase RNA (hTR)⁵, which contains a H/ACA consensus sequence^{6,7}. Here we map the gene responsible for dyskeratosis congenita in a large pedigree with autosomal dominant inheritance. Affected members of this family have an 821-base-pair deletion on chromosome 3q that removes the 3' 74 bases of hTR. Mutations in hTR were found in two other families with autosomal dominant dyskeratosis congenita.

Three other proteins, GAR1, NHP2 and NOP10, are known to be present along with dyskerin in the nucleolar ribonucleoprotein complex and in the telomerase complex^{5,8,9}. Telomerase is an RNA-protein complex that is essential for maintaining the nucleoprotein caps at the ends (telomeres) of eukaryotic chromosomes^{10,11}. The principal components of telomerase are hTR⁶ and a specialized reverse transcriptase (hTERT)¹². Dyskeratosis congenita is a multi-system disease that affects tissues such as skin, gut and bone marrow, all of which require constant renewal that is dependent on stem-cell activity, and thus may be due to a defect in stem-cell turn over or proliferative capacity^{2,13}. Defects in rRNA synthesis and/or in telomere maintenance might affect stem-cell function¹⁴. Dyskeratosis congenita patients have markedly shorter telomeres than normal individuals and this is apparent from an early age¹⁵. The relative importance of rRNA processing and telomere maintenance in the pathophysiology of dyskeratosis congenita may be clarified by the nature of the genetic loci causing the autosomal form(s) of the disease. Our finding of mutations in the telomerase RNA component (hTR) in three separate autosomal dominant pedigrees suggests that dyskeratosis congenita is due to defective telomerase activity.

Among the families on the dyskeratosis congenita registry at the Hammersmith Hospital is a large family from Iowa, USA, with a mild form of dyskeratosis congenita and autosomal dominant inheritance (DCR101; see Supplementary Information for the full

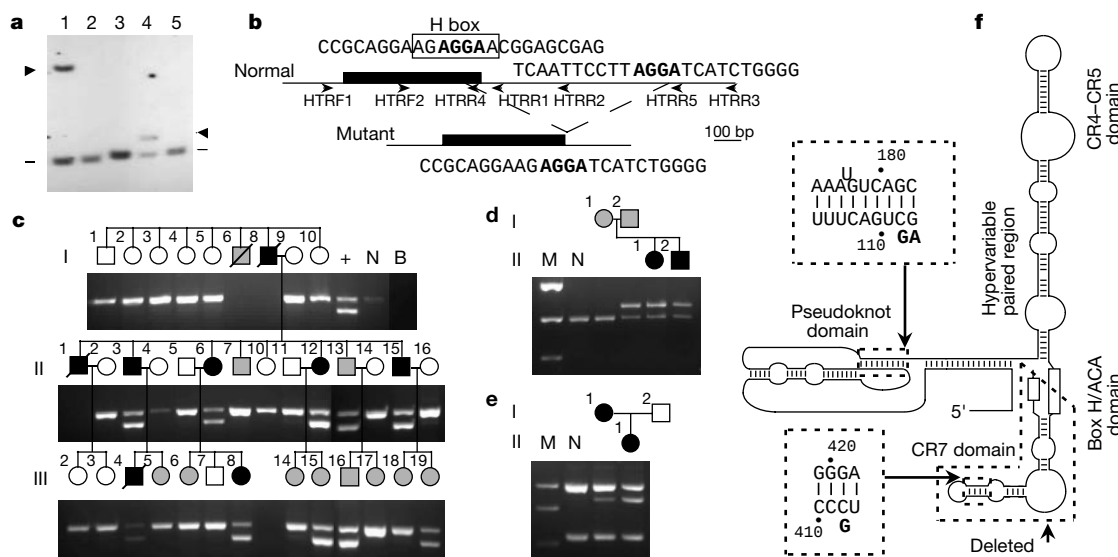


Figure 1 Identification and location of hTR mutations. **a**, Southern blot analysis using an hTR probe to genomic DNA digested with *Pst*I (lanes 1 and 2) and *Taq*I (lanes 3–5) of an affected member from family DCR101 (lanes 1 and 4) and unrelated normal individuals (lanes 2, 3 and 5). Abnormal fragments are indicated by arrowheads; normal fragments are indicated with a dash. **b**, Diagram of an 821-bp deletion that removes the 3' end of the hTR gene. The location, orientation and names of oligonucleotides used in various PCR reactions are given. **c**, Segregation of the deletion. A subset of members of family DCR101 are shown above the appropriate lanes. The larger and smaller fragments are from the normal and mutant alleles, respectively. Shading indicates that individuals were

either too young for diagnosis or showed features that were suggestive but not diagnostic of dyskeratosis congenita. Affected individuals are shown in black; normal individuals in white. Plus, mutant control; N, normal control; B, blank control. **d**, The segregation of an hTR C408G substitution in family DCR063, which results in the loss of a *Tfi*I site. M, molecular mass marker (pEMBL8 plasmid cut with *Taq*I and *Pvu*II). **e**, The presence of a GC to AG double substitution at nucleotides 107–108 of hTR in family DCR082 is confirmed by the creation of a *Dde*I site in affected individuals. **f**, Model of human hTR¹⁶. Dotted boxes show the location of these point mutations and the extent of the deletion observed in family DCR101.

pedigree). In addition to the diagnostic mucocutaneous features, some affected family members had premature greying, early dental loss, bone marrow failure, liver cirrhosis, pulmonary disease and skin cancer (see Supplementary Information). A genome-wide scan was performed using 400 microsatellite markers on DNA from six affected and two unaffected individuals. After the initial stage of the analysis, LOD scores ($\log_{10}$ of the odds ratio in favour of linkage) ranged from -100 to 1.5 with only one region in chromosome 3q showing consistent, adjacent high LOD scores. The addition of other DCR101 family members and more markers narrowed down the region of interest to a segment of approximately 30 centimorgans, the highest LOD score of 1.8 being achieved with the D3S3725 marker. Of all the genes in this region our attention was drawn to the gene encoding the RNA component of telomerase, which maps to 3q21–3q28, as this molecule had already been shown to be associated with dyskerin⁵.

The entire coding region of the gene was amplified from an affected individual (II.6) using a pair of flanking primers. Although the sequence of this fragment was normal, Southern blotting of DNA from affected and normal individuals showed clear differences in the pattern obtained (Fig. 1a), consistent with the existence of a small deletion in the DNA of patients. Further amplification with primers—more distal than the original 3' primer—showed that one allele in the patients' DNA had 821 base pairs (bp) deleted, including the 74 3' base pairs of the coding region (Fig. 1b). The deleted region is flanked by a 4-bp direct repeat (AGGA). Using primers specific for the deletion and for the normal allele, DNA from all members of the pedigree was tested. The results show that all available affected individuals in generation II were heterozygous for the deletion and all unaffected members had two normal alleles (Fig. 1c). Out of the two individuals in generation II (II:7 and II:13) with undetermined clinical status, one (II:13) has the deletion. Individual II:13 was 37 years of age when examined, while affected members of this family from generation II were diagnosed at ages 29, 40, 46 and 48. In generation III all members diagnosed as being affected had the deletion. Some family members were too young to show signs of the disease. Of these, three (III:15, III:16 and III:19) were heterozygous and five (III:5, III:6, III:14, III:17 and III:18) were normal. The three children who were heterozygous were 3, 9 and 13 years of age, whereas the two individuals in this generation who had the disease were diagnosed at 17 and 22 years. The presence of the deletion in these unaffected members (II:13, III:15, III:16 and III:19) reflects the variability in the severity and age of onset in this mild form of dyskeratosis congenita.

Lymphocyte cell lines were derived from affected individuals by transformation with the Epstein Barr Virus (EBV). These cell lines were examined by northern blotting using an hTR probe. No difference was found between cell lines from affected and unaffected individuals (Fig. 2a). All cell lines showed only the normal 451-bp telomerase RNA. Polymerase chain reaction with reverse transcription (RT-PCR) experiments, using a mutant-specific 3' primer abutting the point of the deletion, showed that the mutant transcript is barely detectable (Fig. 2b). Similarly no significant difference was found in the telomerase activity levels measured by a TRAP assay in the EBV cell lines (data not shown). Telomere length in affected family members was measured with a chromosome-7-specific subtelomeric probe. As with our previous observations telomere length in these dyskeratosis congenita patients is significantly shorter than in normal patients¹⁵, even in young children (Fig. 2d).

Having found a lesion in the hTR gene in one family we examined this gene in two other families with autosomal dominant dyskeratosis congenita, as well as in one family with an unknown pattern of inheritance. None of these families showed alteration when analysed with a Southern blot. However, sequence analysis of the hTR gene revealed affected members of one family (DCR063) were heterozygous for a single point mutation at position 408 (C408G).

In another family (DCR082) a double point mutation (GC to AG) at positions 107–108 was observed in affected members. The presence of these mutations and their co-segregation with the disease was confirmed in the respective families by detection of restriction enzyme sites lost (DCR063; loss of *TfiI*, Fig. 1d) or created (DCR082; new *DdeI*, Fig. 1e) by the mutations. In family DCR063 the father (I:2; age 39 years) had only mild haematological abnormalities, whereas his children (II:1 and II:2; 10 and 12 years) had severe bone marrow failure in association with other somatic abnormalities. In family DCR082 both I:1 and II:1 showed severe disease with ages at diagnosis of 52 and 30 years, respectively. The genetic lesions found in all three families were absent in a panel of 50 unrelated individuals that we tested. No abnormality of the hTR gene was detected in the remaining family. This family had a more severe form of the disease than the other three families, with both the father and daughter developing aplastic anaemia by the age of 5 years.

Our results show that in three families, mutations in one allele of the RNA component of telomerase are sufficient to cause dyskeratosis congenita. This could be due to haplo-insufficiency or to a dominant negative effect. The fact that we do not detect an aberrant

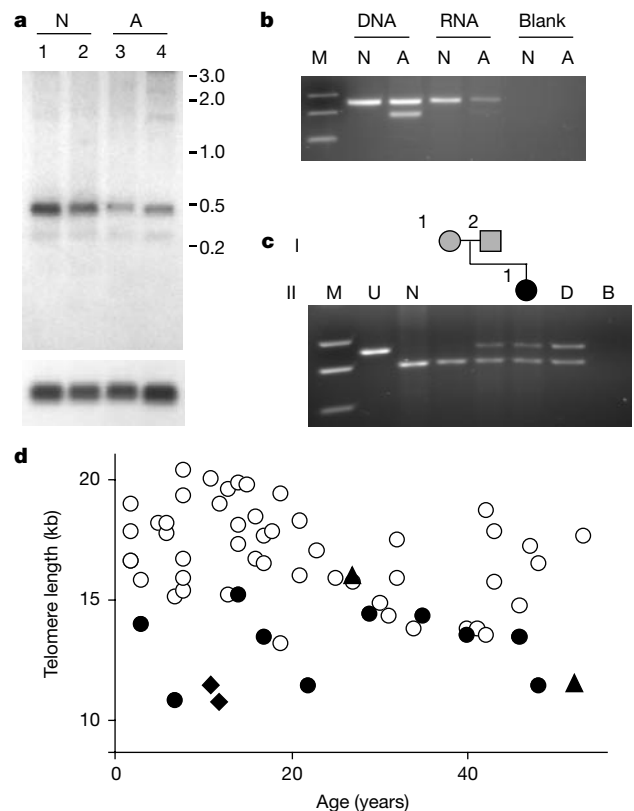


Figure 2 Expression of hTR mutations. **a**, Northern blot analysis of two normal individuals (N, lanes 1 and 2) and two affected members of family DCR101 (A, lanes 3 and 4). The lower panel shows hybridization of a 5S rRNA probe to the same filter. The position and size (kilobases) of molecular mass markers are shown. **b**, PCR and RT-PCR amplification from genomic DNA and primary mononuclear cell RNA as indicated from a normal (N) and an affected (A) individual from family DCR101. No products are seen from cDNA prepared without reverse transcription (blank). **c**, RT-PCR products from EBV lines from members of family DCR063 (drawn above the appropriate lanes) digested with *TfiI*. M, Molecular mass marker; N, normal individual; U, undigested PCR product; D, amplification from genomic DNA of an affected individual; B, amplification from a cDNA sample prepared without reverse transcription. **d**, The telomere lengths of affected individuals from family DCR101 (filled circles), DCR063 (filled diamonds) and DCR086 (filled triangles) are compared with age-matched normal controls (open circles). Telomere lengths are significantly smaller ($P < 0.001$, Student's *t*-test), with short telomeres seen also in the younger, affected individuals.

transcript on northern blots from family DCR101 implicates haplo-insufficiency. The deletion in this family removes the last 74 bp of hTR, which includes the box H/ACA domain and the CR7 domain¹⁶ (Fig. 1f). Mutations in both of these regions have been shown to be essential for telomerase RNA accumulation and 3' end formation^{16,17}. The G408C mutation in family DCR063 would be predicted to destabilize a stem structure in the CR7 domain. Analysis of complementary DNA from two cell lines from this family (Fig. 2c) shows that mutant and wild-type hTRs are present in roughly equal amounts, leaving open the possibility of a dominant negative effect. The double mutation in family DCR082 seems to destabilize a hairpin stem region in the pseudoknot region in the proposed secondary structure of hTR. This region is essential for telomerase activity^{16,17} and seems to be involved in interaction with hTERT¹⁸. Expression studies on this family were not possible owing to a lack of the appropriate material. The mechanism(s) by which hTR mutations lead to the dyskeratosis congenita phenotype will be the focus of future studies.

The effect of mutation of one allele of human telomerase RNA in families with autosomal dominant dyskeratosis congenita can be compared with the effect on laboratory mice of knocking out both alleles of the gene coding for telomerase RNA (mTR)¹⁹. Such mice showed no significant abnormalities in early generations, but later generations showed a variety of defects including reduced proliferative capacity of haematopoietic cells in the bone marrow and spleen^{20,21}. Although the clinical features of dyskeratosis congenita are variable both within and between families, the similarity with mTR^{-/-} mice is evident¹⁴ (see Supplementary Information for a detailed comparison of mTR^{-/-} mice and the patients discussed here). The lag in appearance of the phenotype in mTR^{-/-} mice may well be due to the fact that laboratory strains of mice have relatively long telomeres, and several generations are needed before a critical amount of shortening has occurred²². Wild mice have much shorter telomeres²³, and it will be instructive to determine the effect of mTR ablation on wild strains. In this regard it is interesting to note that in all three families presented here diagnosis of dyskeratosis congenita has been made at a younger age in succeeding generations, although firm conclusions cannot yet be made from this small sample size. Aberrant telomerase activity has been implicated in human cancer as many cancer cells have increased telomerase activity. The finding that both dyskeratosis congenita patients and mTR^{-/-} mice with decreased telomerase expression have an increased incidence of cancer seems paradoxical²⁴. It is likely that in both cases the increased incidence of malignancy is caused by chromosome instability, which in turn is a result of critically shortened telomeres. Indeed, end-to-end chromosome fusions have been observed in both mTR^{-/-} mice²⁵ and in dyskeratosis congenita patients².

Bone marrow failure in X-linked and hTR-related dyskeratosis congenita is due to a defect in telomerase activity. Healthy blood cells would be expected to have a growth advantage over the telomerase-deficient dyskeratosis congenita cells. In female carriers of X-linked dyskeratosis congenita this growth advantage results in 100% skewing of X-inactivation, so only cells expressing the normal allele can be detected²⁶. If the genetic defect could be corrected, or the wild-type gene or gene product introduced into a haematopoietic stem cell, the severe effects of this disease could be improved, thereby making dyskeratosis congenita a good candidate for treatment by the emerging gene transfer methodologies. □

Methods

Linkage analysis

We prepared genomic DNA from whole blood using the Puregene DNA isolation kit (Gentra systems). A genome-wide search for linkage was performed using the LMS-MD10 ABI PRISM set of microsatellite markers (Applied Biosystems). Regions of linkage were refined using additional markers selected from the Genethon map and the LMS-HD5 linkage mapping set. PCR reactions were carried out on 12.5 ng genomic DNA in 1× Gold PCR buffer, 2.5 mM MgCl₂, 1 mM dNTP, 1.25 pmol forward and reverse primer and 0.25 U

Amplitaq (Applied Biosystems). PCR parameters were 95 °C for 12 min, 10 cycles of 95 °C for 15 s, 55 °C for 15 s and 72 °C for 30 s, 20 cycles of 89 °C for 15 s, 55 °C for 15 s and 72 °C for 30 s, followed by a final extension step at 72 °C for 10 min. PCR products were analysed in denaturing polyacrylamide gels on an ABI 377 automated DNA sequencer and scored using the GeneScan and Genotyper analysis software programs (Applied Biosystems). LOD score analysis was performed using the Genetic Linkage User Environment (GLUE, www.hgmp.mrc.ac.uk) assuming 100% penetrance, autosomal dominant inheritance, and rare prevalence (0.001) of the disease allele.

Southern blotting and hTR amplification

Restriction enzyme digestion of human genomic DNA, agarose gel electrophoresis and Southern blotting were carried out using standard methods²⁷. The blots were hybridized to a ³²P-labelled, 653-bp hTR probe, generated by PCR amplification from genomic DNA using the primer pair HTRF1 (5'-TCATGGCCGGAATGGAAGT-3') and HTRR1 (5'-GGGTGACGGATGCGACGAT-3'). Amplification of hTR was performed in the presence of 10% DMSO using the following parameters: 7 min at 95 °C, followed by 30 cycles of 94 °C for 45 s, 58 °C for 45 s and 72 °C for 1 min, followed by extension at 72 °C for 7 min. Amplification of the normal and deleted alleles in family DCR101 was achieved using three primers, HTRF2 (5'-CCTGCCGCTTCCACCGTTCAAT-3'), HTRR2 (5'-AATCTGCAGAGCAGGAAGTAA-3') and HTRR3 (5'-CATTACCAGCAACAGTGGACCTCT-3') in the same reaction. Restriction enzyme digestion of PCR products was performed overnight at the appropriate temperature using commercial buffers (New England Biolabs).

Northern blotting and RT-PCR

RNA was extracted using an RNeasy kit (Qiagen) following the manufacturers recommendations. Northern blotting was performed by the formaldehyde method according to standard procedures²⁷. The probe for 5S RNA was the 41-nucleotide 5'-GGTCTCCATCCAGTACTAACCAGGCCGACCTGCTTAG-3'. The preparation of cDNA from total RNA using random hexamers and DNase 1 was carried out by standard methods. RT-PCR was performed using the primers HTRF2 (see above), HTRR4 (5'-GCATGTGTGAGCCGAGTCCTG-3') and HTRR5 (5'-TCTTGGCAACTACCCCATGATGA-3'). PCR conditions were as above except that 26 cycles were used.

Telomere length

Telomere length was determined by Southern blot analysis using a ³²P-labelled sub-telomeric probe from chromosome 7 (pTelBam8)²⁸ to BamHI digests of genomic DNA extracted from whole blood and run on 0.8% agarose gels²⁹. The size of the peak of signal intensity was determined using the Gel Blot-Pro software from UVP (Upland).

Received 15 March; accepted 11 July 2001.

- Drachtman, R. A. & Alter, B. P. Dyskeratosis congenita: clinical and genetic heterogeneity. Report of a new case and review of the literature. *Am. J. Pediatr. Hematol. Oncol.* **14**, 297–304 (1992).
- Dokal, I. Dyskeratosis congenita in all its forms. *Br. J. Haematol.* **110**, 768–779 (2000).
- Heiss, N. S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genet.* **19**, 32–38 (1998).
- Tollervey, D. & Kiss, T. Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* **9**, 337–342 (1997).
- Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551–555 (1999).
- Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241 (1995).
- Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19**, 567–576 (1999).
- Pogacic, V., Dragon, F. & Filipowicz, W. Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Mol. Cell. Biol.* **20**, 9028–9040 (2000).
- Dragon, F., Pogacic, V. & Filipowicz, W. *In vitro* assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs. *Mol. Cell. Biol.* **20**, 3037–3048 (2000).
- Prescott, J. C. & Blackburn, E. H. Telomerase: Dr Jekyll or Mr Hyde? *Curr. Opin. Genet. Dev.* **9**, 368–373 (1999).
- Collins, K. Mammalian telomeres and telomerase. *Curr. Opin. Cell Biol.* **12**, 378–383 (2000).
- Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
- Alter, B. P. Molecular medicine and bone marrow failure syndromes. *J. Pediatr.* **136**, 275–276 (2000).
- Marciniak, R. A., Johnson, F. B. & Guarente, L. Dyskeratosis congenita, telomeres and human ageing. *Trends Genet.* **16**, 193–195 (2000).
- Vulliamy, T. J., Knight, S. W., Mason, P. J. & Dokal, I. Very short telomeres in the peripheral blood of patients with X-linked and autosomal dyskeratosis congenita. *Blood Cells Mol. Dis.* **27**, 353–357 (2001).
- Chen, J. L., Blasco, M. A. & Greider, C. W. Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–514 (2000).
- Martin-Rivera, L. & Blasco, M. A. Identification of functional domains and dominant negative mutations in vertebrate telomerase RNA using an *in vivo* reconstitution system. *J. Biol. Chem.* **276**, 5856–5865 (2001).
- Bachand, F. & Autexier, C. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. *Mol. Cell. Biol.* **21**, 1888–1897 (2001).
- Blasco, M. A. *et al.* Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
- Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).

22. Herrera, E. *et al.* Disease states associated with telomerase deficiency appear earlier in mice with short telomeres. *EMBO J.* **18**, 2950–2960 (1999).
23. Hemann, M. T. & Greider, C. W. Wild-derived inbred mouse strains have short telomeres. *Nucleic Acids Res.* **28**, 4474–4478 (2000).
24. Artandi, S. E. & DePinho, R. A. A critical role for telomeres in suppressing and facilitating carcinogenesis. *Curr. Opin. Genet. Dev.* **10**, 39–46 (2000).
25. Artandi, S. E. *et al.* Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
26. Vulliamy, T. J., Knight, S. W., Dokal, I. & Mason, P. J. Skewed X-inactivation in carriers of X-linked dyskeratosis congenita. *Blood* **90**, 2213–2216 (1997).
27. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, New York, 1989).
28. Brown, W. R. *et al.* Structure and polymorphism of human telomere-associated DNA. *Cell* **63**, 119–132 (1990).
29. Notaro, R., Cimmino, A., Tabarini, D., Rotoli, B. & Luzzatto, L. *In vivo* telomere dynamics of human hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **94**, 13782–13785 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Alter and E. Gluckman for facilitating recruitment to the dyskeratosis congenita registry, and the dyskeratosis congenita families and their clinicians who provided us with the samples. We also thank Z. Amoura, R. Gruppo and J. Miller. We thank D. Stevens for technical assistance and D. Smyth for help with the linkage. We are grateful to S. Knight, W. Watkins, J. Clarke, J. O'Donnel and S. Reiss for help and discussions, and J. Goldman and L. Luzzatto for their continued support of the dyskeratosis congenita project. The genome screen was performed at the Linkage Hotel MRC UK HGMP Resource Centre. This work was supported by the Wellcome Trust.

Correspondence and requests for materials should be addressed to P.J.M. (e-mail: p.mason1@ic.ac.uk).

DNA topoisomerase II α is required for RNA polymerase II transcription on chromatin templates

Neelima Mondal & Jeffrey D. Parvin

Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

In the nucleus of the cell, core RNA polymerase II (pol II) is associated with a large complex called the pol II holoenzyme (holo-pol)^{1,2}. Transcription by core pol II *in vitro* on nucleosomal templates is repressed compared with that on templates of histone-free naked DNA^{3–5}. We found that the transcriptional activity of holo-pol, in contrast to that of core pol II, is not markedly repressed on chromatin templates. We refer to this property of holo-pol as chromatin-dependent coactivation (CDC). Here we show that DNA topoisomerase II α is associated with the holo-pol and is a required component of CDC. Etoposide and ICRF-193, specific inhibitors of topoisomerase II, blocked transcription on chromatin templates, but did not affect transcription on naked templates. Addition of purified topoisomerase II α reconstituted CDC activity in reactions with core pol II. These findings suggest that transcription on chromatin templates results in the accumulation of superhelical tension, making the relaxation activity of topoisomerase II essential for productive RNA synthesis on nucleosomal DNA.

Biochemically purified core RNA polymerase II (pol II) contains 12 subunits⁶, but, in the nucleus of the cell, pol II is associated with holo-pol^{1,2}. Holo-pol also contains basal transcription factors, suppressors of RNA polymerase B mutations (SRB proteins), chromatin remodelling factors and other regulatory proteins^{7,8}. To identify conditions in which the holo-pol is functionally distinct

from the core pol II enzyme, we compared these factors in transcription under various conditions. All reactions included the purified basal factors, the coactivator PC4 and one of the pol II preparations. Every reaction contained two templates that encoded a 210-nucleotide basal control transcript and a 380-nucleotide transcript, the synthesis of which was stimulated by GAL4-fused activators. Both templates were digested with restriction endonucleases and thus had linear topology.

Core pol II was purified from calf thymus⁹ and holo-pol from a sucrose gradient fraction enriched for this complex¹⁰. These two pol II preparations were normalized by basal polymerase activity on naked templates. Under these conditions, 0.2 μ l (40 ng) of core pol II was equivalent in basal transcription activity to 1.5 μ l of holo-pol (Fig. 1a). In subsequent experiments, these matched amounts of each pol II preparation were used for standard comparison. When comparing these normalized pol II activities for mass of polymerase by western blot, it is apparent that the amount of pol II in the core

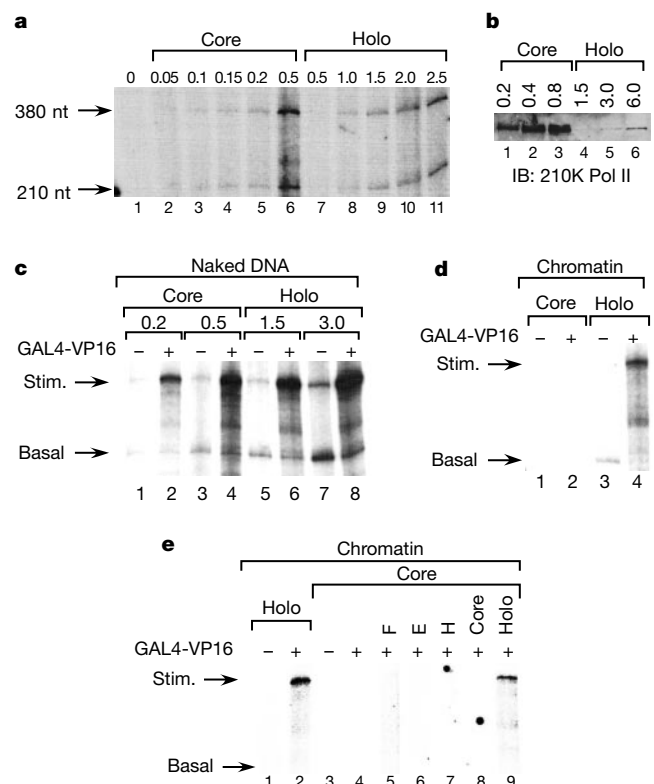


Figure 1 Pol II holoenzyme is qualitatively different from core pol II for transcription on chromatin templates. **a**, Basal transcription reactions with naked DNA templates were used to normalize core pol II (Core) and holo-pol (Holo) preparations. The indicated volumes (in μ l) of each polymerase were added. nt, nucleotide. **b**, Western blot of core pol II and holo-pol. The indicated volumes of core and holoenzyme were subjected to SDS-PAGE and immunoblotted (IB) for the pol II subunit with relative molecular mass 210,000 (M_r 210K). **c**, Activated transcription with naked DNA templates. The volumes (in μ l) of core pol II and holo-pol included in each reaction were varied as indicated. The transcriptional activator GAL4-VP16 was included in half of the reactions, as indicated. The 380-nucleotide GAL4-responsive transcript (Stim.) and the 210-nucleotide basal control transcript (Basal) are indicated. **d**, Transcription of reconstituted chromatin templates. Core pol II (40 ng) and holo-pol (1.5 μ l), matched for basal activity as in **a**, were assayed for transcription on chromatin templates in the absence or presence of the activator GAL4-VP16. **e**, The CDC activity is not due to a basal factor. Reactions with chromatin templates were supplemented by doubling the amount of the indicated factors. Eight nanograms of TFIIE (E, lane 5), 200 ng of TFIIF (F, lane 6), 1 μ l of TFIIH (H, lane 7) and 80 ng of core pol II (lane 8) were included in the indicated reactions. Along with 40 ng of core pol II, 1.5 μ l of holoenzyme was included in lane 9. GAL4-VP16 was included as indicated.

preparation far exceeds that in the holo-pol preparation yielding similar levels of basal transcription (Fig. 1b). For the transcription assay activated by GAL4-VP16 on naked templates, holo-pol was only marginally more responsive than was core pol II (Fig. 1c), which is consistent with our prior observations¹¹. We next tested the same linear DNA templates, reconstituted in chromatin using purified histone octamers¹². Transcription by core pol II was repressed, which is consistent with prior observations^{3–5}. In contrast, holo-pol was active, and the major transcript was encoded by the GAL4-VP16-stimulated template (Fig. 1d). This qualitative difference between core pol II and holo-pol we named chromatin-dependent coactivation (CDC).

Along with core pol II, the basal transcription factors TFIIF, TFIIE and TFIIH are present in the holo-pol preparation^{13,14}. We tested whether an additional amount of basal transcription factors could account for the CDC seen with holo-pol. Excess of these factors had no stimulatory effect on the transcription from chromatin templates, but, when holo-pol was included, RNA was synthesized (Fig. 1e). This result suggests that the CDC regulation requires a novel activity in the holo-pol preparation.

Holo-pol relaxed supercoiled DNA in the presence of ATP, indicating the presence of a DNA topoisomerase II activity in the holo-pol preparation. The activity of DNA topoisomerase II was shown to be stably associated with the holo-pol complex by immunoprecipitation from this protein fraction using affinity-purified antibody specific for SRB7, a holo-pol-specific polypeptide¹¹. As a specificity control, antigenic peptide blocked immunoprecipitation of topoisomerase II but a control peptide did not (Fig. 2a). Immunoprecipitation from the holo-pol preparation

using the SRB7 antibody revealed efficient purification of topoisomerase II α (Fig. 2b). Similar immunoblots specific for topoisomerase II β were negative (data not shown).

The holo-pol-associated relaxation activity of topoisomerase II α was inhibited by 60 μ M etoposide (Fig. 2c). Etoposide also inhibited the CDC activity. GAL4-VP16-stimulated transcription with holo-pol on chromatin templates was inhibited by 60 μ M etoposide. In contrast, when the same transcription reactions were repeated using naked DNA templates, 60 μ M etoposide did not inhibit transcription by either holo-pol or core pol II (Fig. 2d). A second specific inhibitor of topoisomerase II α , ICRF-193, can bind topoisomerase in the absence of DNA in the closed clamp conformation¹⁵. We preincubated the holo-pol with 50 μ M ICRF-193 and then assembled transcription reactions. ICRF-193 inhibited transcription of the chromatin template but had no effect on transcription of the naked template (Fig. 2e). These results suggest that the inhibition of transcription on chromatin templates is not caused by an inhibitor–topoisomerase–DNA complex that nonspecifically blocks elongation. This result demonstrates that the DNA topoisomerase II α is dispensable for RNA synthesis on naked templates but is required for productive transcription on chromatin templates.

We next tested whether topoisomerase II α is sufficient for CDC using core pol II and other defined transcription factors. Purified DNA topoisomerase II α was included at the time of assembling the transcription reaction or while setting up the dialysis for chromatin assembly. The topoisomerase did not block chromatin formation because these templates were still resistant to micrococcal nuclease (see Supplementary Information). Addition of topoisomerase II α

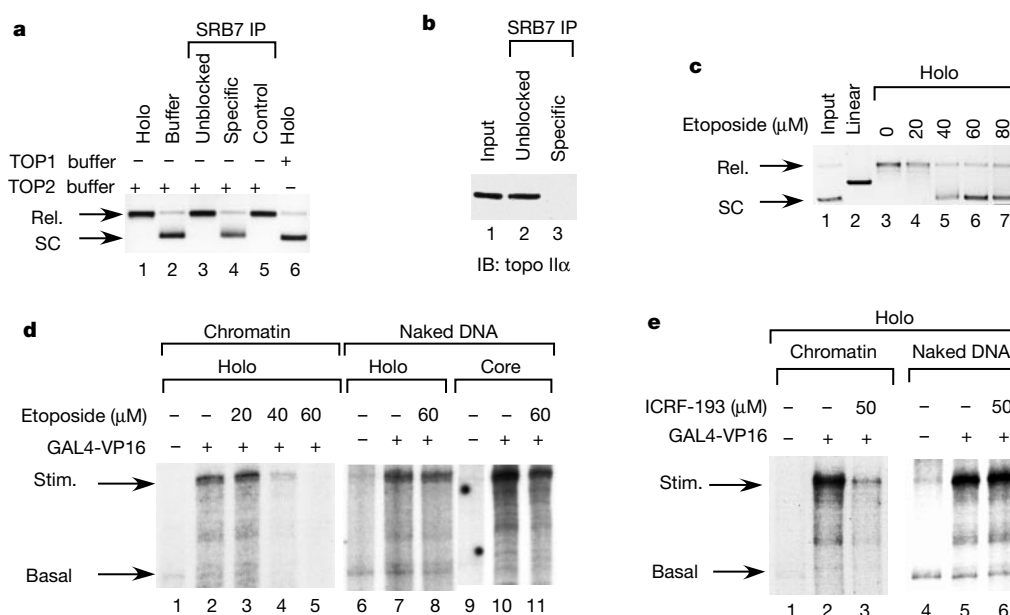


Figure 2 Topoisomerase II α associated with holo-pol is required for CDC. **a**, Holo-pol was assayed for topoisomerase activities using ATP-containing topoisomerase II (TOP2) buffer or ATP-independent topoisomerase I (TOP1) buffer in the indicated reactions. Holo-pol (Holo, lanes 1 and 6) or immunoprecipitate (IP; lanes 3–5) from the same fractions were analysed for relaxation activity. Anti-SRB7 IP was unblocked (lane 3), blocked with 30 μ g of specific antigenic peptide (lane 4) or blocked with 30 μ g of control peptide (lane 5). The image is the negative of a photograph of a gel stained with ethidium bromide. Topoisomerase activity is present if the supercoiled plasmid (SC) is converted to the relaxed form (Rel.). **b**, The holo-pol preparation was tested by immunoblotting (IB) for the presence of DNA topoisomerase II α (topo II α) by IP using affinity-purified anti-SRB7 antibody. Samples were 10% input (lane 1), unblocked IP (lane 2) and IP blocked with the peptide against which the antibody was generated (lane 3). The indicated band migrated at a position consistent with a polypeptide with M_r 170K. **c**, Relaxation activity associated

with holo-pol was inhibited by etoposide. Relaxation assays containing holo-pol (lanes 3–7) were tested in the presence of 20–80 μ M of etoposide (lanes 4–7). The image is negative of a photograph of an ethidium-bromide gel. Inhibition results in the retention of supercoiled plasmid. **d**, Etoposide inhibits transcription by holo-pol on chromatin templates. In the presence of different concentrations of etoposide, transcription reactions used chromatin (lanes 1–5) or naked (lanes 6–11) templates. Holo-pol, core pol II (Core), GAL4-VP16 and etoposide were included in reactions as shown. The 380-nucleotide GAL4-responsive transcript (Stim.) and the 210-nucleotide basal control transcript (Basal) are indicated. **e**, ICRF-193 inhibits transcription by holo-pol on chromatin templates. Holo-pol was preincubated with 100 μ M ATP (lanes 1, 2, 4 and 5) or 100 μ M ATP plus 50 μ M ICRF-193 (lanes 3 and 6) for 30 min at 30 °C. Transcription reactions containing chromatin (lanes 1–3) or naked DNA (lanes 4–6) were processed as usual.

enzyme before chromatin assembly reconstituted the CDC activity in reactions containing core pol II (Fig. 3a). In contrast, addition of topoisomerase enzyme at the time of transcription set-up did not rescue the activated transcription (data not shown).

When adding DNA topoisomerase II α with holo-pol to transcription reactions, the topoisomerase is required to be present only during the transcription reaction and not during chromatin assembly (Fig. 1d). In contrast, addition of purified topoisomerase directly to transcription reactions with core pol II could not reconstitute CDC, but instead the topoisomerase had to be added to the templates before chromatin assembly (Fig. 3a). We therefore

tested whether the purified topoisomerase in the latter case was required during the transcription stage of the reaction. We reconstituted chromatin in the presence of topoisomerase II α and tested whether the addition of etoposide at the time of transcription could block CDC (Fig. 3b). We found that topoisomerase II α was required during the transcription phase of the reaction because 80 μ M etoposide added at the time of transcription inhibited the core pol II reaction (Fig. 3b). As before, under these conditions, the activity of core pol II on naked templates was not inhibited by etoposide (Fig. 3c), indicating that etoposide was not nonspecifically blocking transcription. Transcription dependent upon purified topoisomerase II α and core pol II on chromatin was inhibited by ICRF-193 whereas transcription on naked templates was unaffected (Fig. 3d). These results suggest that the binding of topoisomerase II α to DNA is blocked by preformed chromatin.

The results in Fig. 3a–d demonstrate that topoisomerase II α can function in transcription on chromatin templates, but the timing of addition suggests that the holo-pol preparation contains a second activity that may facilitate the interaction of the topoisomerase II α with the chromatin-bound DNA. We tested for a second activity by disrupting the holo-pol using an antibody against the carboxy-terminal domain of the largest subunit of pol II and then depleted the protein preparation of pol II by use of protein A beads^{1,11}. The supernatant proteins contained topoisomerase II α (data not shown). This supernatant was added to transcription reactions containing core pol II and chromatin templates, and transcripts accumulated dependent on the presence of the disrupted holoenzyme proteins and core pol II. Transcription under these conditions was inhibited by 50 μ M ICRF-193, consistent with the requirement for topoisomerase II α (Fig. 3e). These results suggest that, along with topoisomerase II α , some other factor is required for transcription from preformed chromatin templates. This additional factor is dispensable if topoisomerase is added before chromatin assembly, suggesting that the additional activity is loading the topoisomerase on to the template.

Because topoisomerase II α is associated with the holo-pol, and because two different, highly specific inhibitors of topoisomerase II inhibit the CDC, and because purified topoisomerase II α can reconstitute the CDC in a defined transcription system, it is highly likely that it is topoisomerase II α in the holo-pol that mediates CDC. A role for topoisomerase II in transcription has long been noted¹⁶, and in these experiments we have identified a mechanism by which the topoisomerase can regulate transcription.

We propose a model (Fig. 4) that is based on established phenomena related to superhelical tension and transcription¹⁷. The transcription process generates supercoils on the template: positive supercoils ahead of the elongating pol II and negative supercoils in its wake. The positive superhelical tension generated by RNA synthesis hyperwinds the DNA and inhibits transcription^{18,19}. On naked DNA templates with free ends (linear

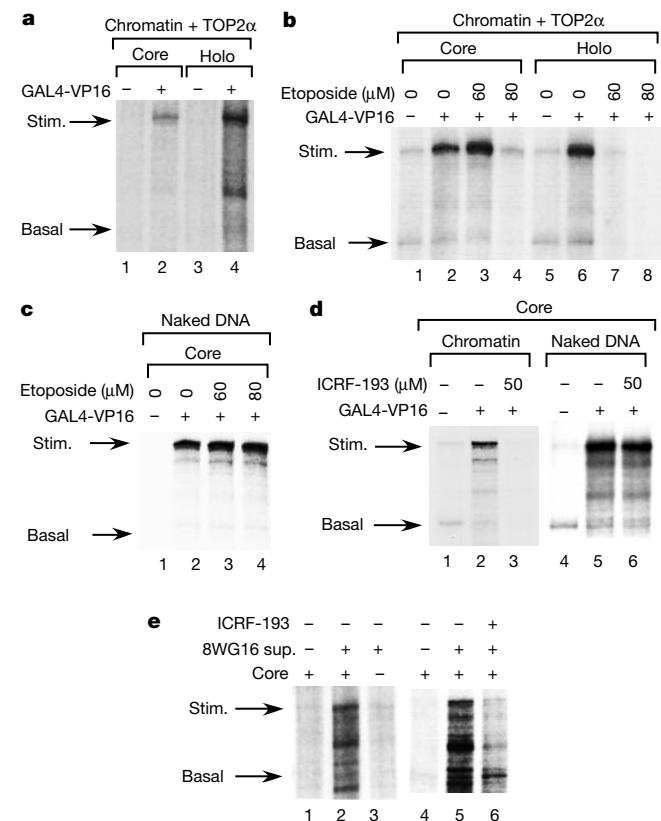


Figure 3 Topoisomerase II α rescues transcriptional activity on chromatin by core pol II. **a**, Transcription reactions dependent upon core pol II (Core, lanes 1 and 2) or holo-pol (Holo, lanes 3 and 4) were tested in the presence of GAL4-VP16 activator on chromatin templates reconstituted in the presence of purified topoisomerase II α (TOP2 α). The 380-nucleotide GAL4-responsive transcript (Stim.) and the 210-nucleotide basal control transcript (Basal) are indicated. **b**, CDC activity of topoisomerase II α with core pol II is inhibited by etoposide. Chromatin was assembled in presence of topoisomerase II α . The templates were subjected to *in vitro* transcription either by core pol II (lanes 1–4) or holo-pol (lanes 5–8). Gal4-VP16 and etoposide were included as indicated. **c**, Transcription reactions as in **b** lanes 1–4 were repeated using naked DNA templates. **d**, CDC activity of topoisomerase II α is inhibited by ICRF-193. Templates of chromatin plus topoisomerase II α (lanes 1–3, as in **a** and **b**) or of naked DNA (lanes 4–6) were preincubated with either 100 μ M ATP (lanes 1, 2, 4 and 5) or 100 μ M ATP plus 50 μ M ICRF-193 (lanes 3 and 6) for 30 min. Transcription reactions were then processed as usual. **e**, A second activity associated with the holo-pol, along with topoisomerase II α , is required for transcription on preformed chromatin. The holo-pol preparation was treated with 8WG16 antibody, which disrupts the holo-pol by binding to the largest pol II subunit, and antibody and core pol II were removed from the preparation using protein A agarose. The 8WG16 supernatant (sup.) was added (lanes 2, 3, 5 and 6) to preformed chromatin, along with core pol II (lanes 1, 2 and 4–6). Transcription reactions in lanes 1–3 were immediately assembled and processed as usual. Reactions in lanes 4–6 were preincubated for 30 min with 100 μ M ATP (lanes 4 and 5) or with 100 μ M ATP plus 50 μ M ICRF-193 (lane 6) before addition of the transcription factors and were processed as usual. GAL4-VP16 was present in all reactions.

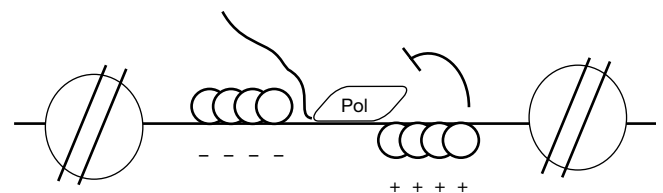


Figure 4 Model for requirement of topoisomerase II α activity in pol II transcription. We observed a requirement for topoisomerase II α in transcription, even though the DNA had linear topology. We propose that nucleosomes provide a barrier against diffusion of supercoils that result from the polymerization process. Positive supercoils, which accumulate ahead of the elongating pol II, hyperwind the DNA and block further elongation. According to this model, topoisomerase II α relieves the elongation block by relaxing the template DNA.

topology), the superhelical tension generated by the transcription process diffuses off the free ends. By contrast, our data suggest that in chromatin the diffusion of superhelical turns is blocked even though there are free DNA ends. Thus, nucleosomes topologically close the system, and accumulated positive superhelical tension would inhibit elongation. DNA topoisomerase II α , which could relieve the superhelical tension, rescued transcription on these templates. Although we have demonstrated that topoisomerase II α has a required function in transcription on chromatin and that chemicals which inhibit the relaxation activity of the topoisomerase II α inhibit the transcription activity, it is formally possible that the transcription function we have observed is dependent on some as-yet-unidentified activity of topoisomerase II α . Future experiments on the mutagenesis of topoisomerase II α will clarify this point.

In yeast, the inactivation of temperature-sensitive mutants of both topoisomerase I and II results in inhibition of transcription by RNA pol I (refs 19, 20) and a threefold decrease in transcription by RNA pol II, although three specific genes analysed were unaffected by loss of topoisomerase²⁰. These genetic data support the direct role for topoisomerases in transcription and are analogous to what we have demonstrated *in vitro* for pol II using purified factors. □

Methods

Plasmid templates

G-less cassette construct G5-E4 (ref. 21) contained GAL4 response elements, the adenoviral E4 promoter and a sequence containing no guanines in the coding strand from the transcription start site to 380 base pairs downstream. A basal control template, pAML-200, encoded a 210-base-pair G-less cassette. Both templates were digested with *Xmn*I, thus these DNAs have linear topology.

Chromatin assembly

Histones were purified from the chromatin pellet of HeLa whole-cell extracts²² with hydroxyapatite matrix¹². For some experiments, the histone octomers were further purified by gel filtration and the elution was consistent for the predicted relative molecular mass of 110,000 for histone octomers. Purified histone octomers and the linear DNA templates were mixed in the ratio of 0.8 mg of histone to 1 mg of total DNA. Typically, 8.4 μ g of each DNA template was mixed with 13.5 μ g of purified octomers. Histones were bound to the templates by slow dialysis¹².

Purified DNA topoisomerase II α was obtained from Topogen. One unit of purified DNA topoisomerase II α per 0.25 μ g of DNA was used for the reconstitution of chromatin, wherever indicated.

Purification of transcription factors

The transcription factors used in this study were purified by established techniques. Holo-pol from BJA-B whole-cell extract was purified by Biorex 70 chromatography and sucrose gradient sedimentation¹⁰. Core pol II was purified from calf thymus by DEAE chromatography and immuno-affinity purification with antibody 8WG16 (ref. 9). TFIIF was purified from HeLa whole-cell extracts by chromatography using the following matrices: Biorex 70, DEAE sepharose, bioscale S and bioscale Q²³. Flag-tagged TFIID was purified from HeLa cells and PC4 and TFIIA, -B, -F and -E were expressed and purified from bacteria^{23–25}.

Transcription reactions were performed using published techniques²⁵.

Relaxation assay

Antisera specific for topoisomerase II α and II β were obtained from PharMingen and BioTrend Chemicals, respectively. Relaxation assays contained plasmid DNA (200 μ g per reaction) in a final volume of 30 μ l in either topoisomerase II buffer (50 mM Tris-Cl at pH 8.0, 120 mM KCl, 10 mM MgCl₂, 0.5 mM ATP, 0.5 mM DTT, 30 μ g ml⁻¹ BSA) or topoisomerase I buffer (10 mM Tris-Cl at pH 7.9, 1 mM EDTA, 150 mM NaCl, 0.1% BSA, 0.1 mM spermidine, 0.5% glycerol). The reactions were incubated at 37 °C for 30 min, terminated by the addition of 2.5 μ l bromophenol blue and phenol/chloroform

extraction, and then analysed on 1% native agarose gel without intercalator. Gels were stained with ethidium bromide.

Received 22 May; accepted 1 August 2001.

- Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* **77**, 599–608 (1994).
- Koleske, A. J. & Young, R. A. An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466–469 (1994).
- Laybourn, P. J. & Kadonaga, J. T. Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. *Science* **254**, 238–245 (1991).
- LeRoy, G., Orphanides, G., Lane, W. S. & Reinberg, D. Requirement of RSF and FACT for transcription of chromatin templates *in vitro*. *Science* **282**, 1900–1904 (1998).
- Sandaltzopoulos, R., Blank, T. & Becker, P. B. Transcriptional repression by nucleosomes but not H1 in reconstituted preblastoderm *Drosophila* chromatin. *EMBO J.* **13**, 373–379 (1994).
- Cramer, P. et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649 (2000).
- Cho, H. et al. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol. Cell. Biol.* **18**, 5355–5363 (1998).
- Parvin, J. D. & Young, R. A. Regulatory targets in the RNA polymerase II holoenzyme. *Curr. Opin. Genet. Dev.* **8**, 565–570 (1998).
- Thompson, N. E., Aronson, D. B. & Burgess, R. R. Purification of eukaryotic RNA polymerase II by immunoaffinity chromatography: Elution of active enzyme with protein stabilizing agents from a polyclonal-responsive monoclonal antibody. *J. Biol. Chem.* **265**, 7069–7077 (1990).
- Neish, A. S., Anderson, S. F., Schlegel, B. P., Wei, W. & Parvin, J. D. Factors associated with the mammalian RNA polymerase II holoenzyme. *Nucleic Acids Res.* **26**, 847–853 (1998).
- Chao, D. M. et al. A mammalian SRB protein associated with an RNA polymerase II holoenzyme. *Nature* **380**, 82–85 (1996).
- Workman, J. L., Taylor, I. C., Kingston, R. E. & Roeder, R. G. Control of class II gene transcription during *in vitro* nucleosome assembly. *Methods Cell Biol.* **35**, 419–447 (1991).
- Cho, H., Maldonado, E. & Reinberg, D. Affinity purification of a human RNA polymerase II complex using monoclonal antibodies against transcription factor IIF. *J. Biol. Chem.* **272**, 11495–11502 (1997).
- Scully, R. et al. BRCA1 is a component of the RNA polymerase II holoenzyme. *Proc. Natl Acad. Sci. USA* **94**, 5605–5610 (1997).
- Ishida, R. et al. DNA topoisomerase II is the molecular target of bisdioxopiperazine derivatives ICRF-159 and ICRF-193 in *Saccharomyces cerevisiae*. *Cancer Res.* **55**, 2299–2303 (1995).
- Liu, L. F. DNA topoisomerase poisons as antitumor drugs. *Annu. Rev. Biochem.* **58**, 351–375 (1989).
- Wang, J. C. & Lynch, A. S. Transcription and DNA supercoiling. *Curr. Opin. Genet. Dev.* **3**, 764–768 (1993).
- Gartenberg, M. R. & Wang, J. C. Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proc. Natl Acad. Sci. USA* **89**, 11461–11465 (1992).
- Schultz, M. C., Brill, S. J., Ju, Q., Sternglanz, R. & Reeder, R. H. Topoisomerases and yeast rRNA transcription: Negative supercoiling stimulates initiation and topoisomerase activity is required for elongation. *Genes Dev.* **6**, 1332–1341 (1992).
- Brill, S. J., DiNardo, S., Voelkel-Meiman, K. & Sternglanz, R. Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature* **326**, 414–416 (1987); erratum *Nature* **326**, 812 (1987).
- Chasman, D. I., Leatherwood, J., Carey, M., Ptashne, M. & Kornberg, R. D. Activation of yeast polymerase II transcription by herpesvirus VP16 and GAL4 derivatives *in vitro*. *Mol. Cell. Biol.* **9**, 4746–4749 (1989).
- Manley, J. L., Fire, A., Samuels, M. & Sharp, P. A. *In vitro* transcription: Whole-cell extract. *Methods Enzymol.* **101**, 568–582 (1983).
- Haile, D. T. & Parvin, J. D. Activation of transcription *in vitro* by the BRCA1 carboxyl-terminal domain. *J. Biol. Chem.* **274**, 2113–2117 (1999).
- Chiang, C. M. & Roeder, R. G. Expression and purification of general transcription factors by FLAG epitope-tagging and peptide elution. *Pept. Res.* **6**, 62–64 (1993).
- Schlegel, B. P., Green, V. J., Ladas, J. A. & Parvin, J. D. BRCA1 interaction with RNA polymerase II reveals a role for hRPB2 and hRPB10 α in activated transcription. *Proc. Natl Acad. Sci. USA* **97**, 3148–3153 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Dutta, J. Wohlschlagel, Z. Jonsson and S. K. Dhar for encouragement and advice throughout the course of this study; U. Banerjee, M. Kannapiran and W. Wei for preparation of extracts; and J. C. Wang for the gift of ICRF-193. This work was supported by a grant from the National Institutes of Health.

Correspondence and requests for materials should be addressed to J.D.P. (e-mail: jparvin@rics.bwh.harvard.edu).

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andrew Douglas (4975)

Holland/ Italy: Nevin Bayouni (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Welss (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@nature.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

MG Ichigaya Building (6F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@nature.jp

Japan Manager:

Kate Cowan

naturejobs

nature.com

Rewards of seeing potential

Youth and enthusiasm: those two characteristics dominated my impressions during a visit to the Sanger Centre in Hinxton, near Cambridge, UK, earlier this month. They are qualities that the centre has used to make up for shortages in trained bioinformatics and sequencing technicians. Rather than cast a global recruiting net, the centre chose to look more locally. Rather than focus on PhD pedigree, recruiters have emphasized potential.

Many of the centre's home-grown talent came to the Sanger with little formal science training, often nothing beyond the odd university course. Although potential is less tangible than a degree, the centre seems to have successfully tapped into it, as recognized by these people's contributions to sequencing the human genome and other organisms.

Training people and recognizing them for their efforts — technicians at the Sanger routinely find their names on scientific papers — pays off, says Carol Churcher, a group leader for pathogen sequencing. Churcher, who herself has no postgraduate degree, says that technicians tend to stay for several years; seldom do they leave within months. Some of them, with the centre's assistance, go on to more formal training.

As more projects emphasize discovery science over the hypothesis-driven approach, there could be room for more people keen to contribute who, for one reason or another, never pursued a postgraduate degree. In order to attract these people, research institutions must be willing to train them and to treat them as partners rather than as service employees. The Sanger Centre is already recognized as a model of genomic excellence. Perhaps it should also be considered as a model of scientific training.

Paul Smaglik

Naturejobs editor



Contents

**WWW.NATURE.COM/
NATUREJOBS**

Career centre
Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS